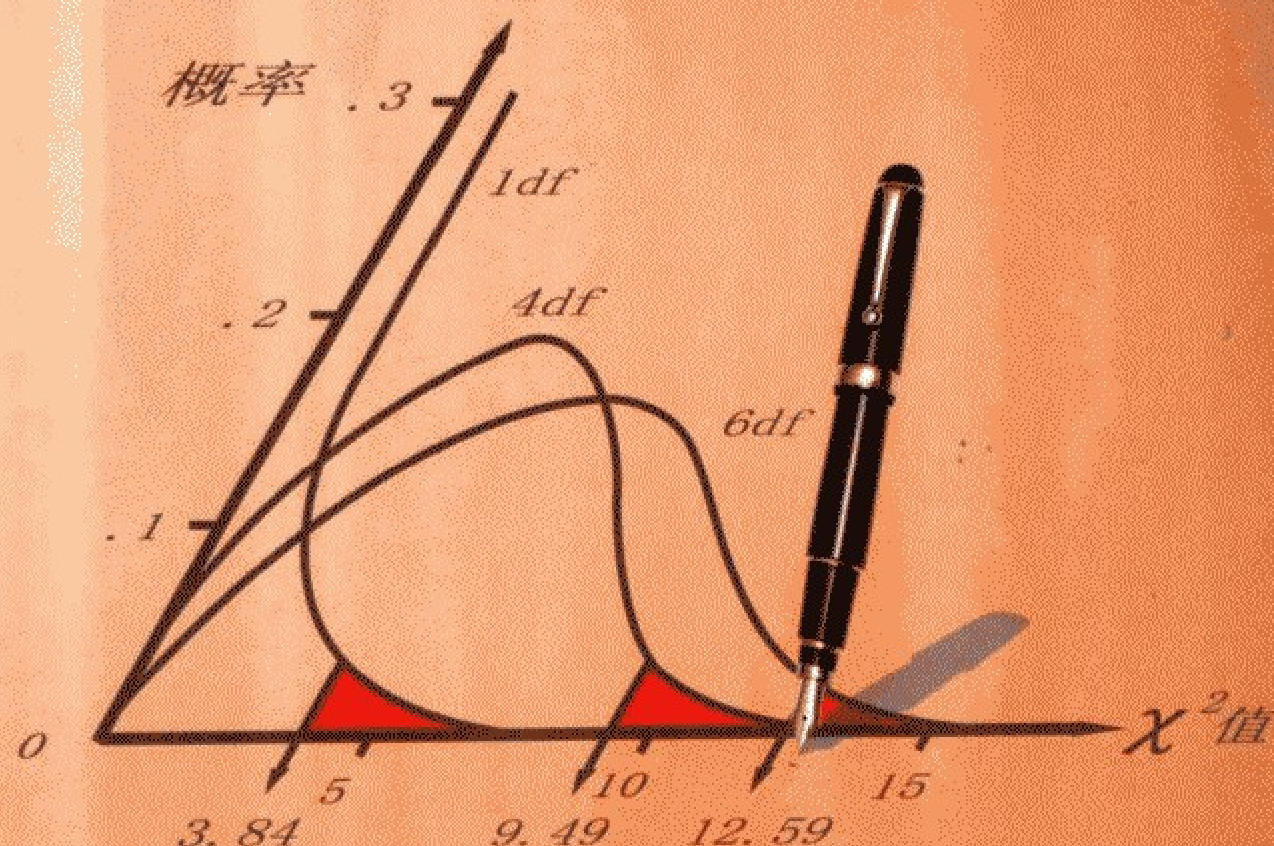
TURING 图灵数学·统计学丛书

WILEY



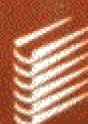
Practical Nonparametric Statistics

实用非参数统计

（第3版）

[美] W.J.Conover 著

崔恒建 译



人民邮电出版社
POSTS & TELECOM PRESS

VISIT…

LANZAROTE

Caliente.com

TURING 图灵数学·统计学丛书

Practical Nonparametric Statistics

实用非参数统计

(第 3 版)

[美] W. J. Conover 著
崔恒建 译

人民邮电出版社
POSTS & TELECOM PRESS

图书在版编目（CIP）数据

实用非参数统计：第3版 /（美）康诺华著；崔恒建译.
—北京：人民邮电出版社，2006.4
（图灵数学·统计学丛书）
ISBN 7-115-14616-0

Ⅰ. 实… Ⅱ. ①康…②崔… Ⅲ. 非参数统计 Ⅳ. 0212.7

中国版本图书馆 CIP 数据核字（2006）第 022972 号

内 容 提 要

非参数统计为有效地分析试验设计及其实际问题中所获得的数据提供了丰富而有说服力的统计工具. 而本书则从问题背景与动机、方法引进、理论基础、计算机实现、应用实例、文献综述等诸多方面介绍了非参数统计方法，其内容包括：基于二项分布的检验、列联表、秩检验、Kolmogorov-Smirnov 型统计量等. 本书内容丰富、思路清晰、层次分明. 在强调实用性的同时，突出了应用方法与理论的结合. 书中的正文和习题中都提供了大量的实际案例，书中最后有许多统计用表以及奇数号习题解答和术语索引.

本书作为非参数统计的基础教材，适用于统计学专业的高年级本科生和研究生课程的教学，也可提供给从事统计学应用与研究、数据的分析处理及其相关领域的专业人员阅读与参考.

图灵数学·统计学丛书

实用非参数统计（第3版）

◆ 著　［美］ W. J. Conover
　译　崔恒建
　责任编辑　王丽萍　蔡红艳

◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街 14 号
　邮编　100061　电子函件　315@ptpress.com.cn
　网址　http://www.ptpress.com.cn
　北京铭成印刷有限公司印刷
　新华书店总店北京发行所经销

◆ 开本：700×1000　1/16
　印张：26.75
　字数：540 千字　2006 年 4 月第 1 版
　印数：1 – 4 000 册　2006 年 4 月北京第 1 次印刷

著作权合同登记号　图字：01-2005-5222 号

ISBN 7-115-14616-0/TP · 5307

定价：59.00 元

读者服务热线：(010)88593802　印装质量热线：(010)67129223

目　录

第 1 章　概率论 …… 1
导言 …… 1
1.1　计数 …… 1
1.2　概率 …… 7
1.3　随机变量 …… 14
1.4　随机变量的性质 …… 21
1.5　连续型随机变量 …… 34
1.6　第 1 章复习题 …… 43
第 2 章　统计推断 …… 46
导言 …… 46
2.1　总体、样本与统计量 …… 46
2.2　估计 …… 54
2.3　假设检验 …… 66
2.4　假设检验的性质 …… 74
2.5　非参数统计评述 …… 81
2.6　第 2 章复习题 …… 85
第 3 章　基于二项分布的检验 …… 88
导言 …… 88
3.1　二项检验与 p 的估计 …… 88
3.2　分位数检验和 χ_p 的估计 …… 97
3.3　容忍限 …… 107
3.4　符号检验 …… 112
3.5　符号检验的一些变形 …… 118
第 4 章　列联表 …… 128
导言 …… 128
4.1　2×2 列联表 …… 128
4.2　$r \times c$ 列联表 …… 142
4.3　中位数检验 …… 156
4.4　相依性度量 …… 163
4.5　χ^2 拟合优度检验 …… 172
4.6　相关观测的 Cochran 检验 …… 180
4.7　其他分析方法讨论 …… 186
4.8　第 3、4 章复习题 …… 187
第 5 章　秩检验 …… 194
导言 …… 194
5.1　两个独立样本 …… 195
5.2　多个独立样本 …… 207
5.3　等方差检验 …… 216
5.4　秩相关性度量 …… 225
5.5　非参数线性回归方法 …… 240
5.6　单调回归方法 …… 250
5.7　一样本或配对情形 …… 255
5.8　多个相关的样本 …… 268
5.9　平衡的不完全区组设计 …… 283
5.10　A. R. E. 不低于 1 的检验 …… 290
5.11　Fisher 随机化方法 …… 299
5.12　秩变换的讨论 …… 307
5.13　第 5 章复习题 …… 310
第 6 章　Kolmogorov-Smirnov 型统计量 …… 317
导言 …… 317
6.1　Kolmogorov 拟合优度检验 …… 317
6.2　分布族的拟合优度检验 …… 327
6.3　两组独立样本的检验 …… 338
6.4　第 1 章至第 6 章复习题 …… 346
附表 …… 353
奇数号习题解答 …… 399
索引 …… 402
参考文献[1]

1. 限于篇幅，本书中参考文献请读者登录图灵网站 www.turingbook.com 查阅下载.

第1章 概 率 论

导 言

非参数统计方法中一个诱人的特性是：你并不需要成为一个概率论方面的专家就能理解非参数方法所蕴含的理论. 只要掌握了一些易学的基本概念，非参数方法的基本理论很容易被理解. 本章介绍这些基本概念，所需要的只是耐心、信心和较好的高中代数知识.

这本书是这样安排的，读者可以直接找到所需要的统计方法，然后从头到尾一步一步按着说明操作. 然而如果这样的话，不一定能知其然更不能知其所以然，从而通常导致错误处理了数据和得出不合理的结论. 在第1、2章，读者应该全面理解所使用的非参数方法，甚至能对其稍作修改，使它能更好地应用于所分析的特殊数据集.

对学习每一节的建议：通读教材，然后演算例子，最后做每一节后的习题和思考题. 这样可以为下一节的学习做准备，也可以增强前面所提到的耐心和信心.

1.1 计 数

计算概率的过程通常依赖于计数，如通常的计数，“1、2、3”等等. 而通常的 5
计数方法在一些复杂的情形下将变得十分冗长乏味，所以本节介绍一些有技巧性的计数方法，用于处理这种复杂的情形.

掷一枚硬币时，我们仅考虑2种可能的结果：或者出现正面（H），或者出现反面（T）. 如果掷一次，可能出现2种结果：H或T；如果掷两次，会出现4种可能：HH,HT,TH,TT，其中HT表示掷第1次时出现H，掷第2次时出现T. 每多掷一次，则可能出现的结果数将是原来的2倍，因为最后一次投掷总有2种可能的结果. 所以，如果投掷n次硬币，结果会出现2^n种可能.

试验

为讨论一般化，我们把掷硬币作为试验的一个例子. 无论是掷一次、两次还是n次，这个过程都被看作是一个试验，这样投掷3次硬币也是一次试验，它是3个仅掷一次硬币的独立试验的组合. 我们称较短的试验为基本试验，“试验”由基本试验组合而成. 一般来说，一个试验是遵循一套已设计好的规则的过程，试验前并不知道

遵循这些规则的最终结果.

模型

很少人会认真地考虑掷硬币作为一个试验其自身所具有的价值. 实际上，掷硬币的价值在于它为许多不同情况下的不同模型提供了原型. 如果我们考虑的是一个均匀硬币，即每一面出现的可能性是一样的，这个试验就与下面的试验类似：老鼠对于两个门的选择；顾客对两种商品的选择；教师判断两种教学法方法哪种更有效；市场分析家预测周一行情将是看涨还是走跌等等.

如果硬币是不均匀的，即一面比另一面更容易朝上，这种模型将适用于更广泛的一类试验. 例如：给老鼠的血液注射药物来检验这种药物是否是致命的；找病人来检验一种新的治疗法；消费者从几种产品里选购一种，这些产品中只有一种是由某公司制造的等等. 每种情况我们关心的结果只有两个，如“生”或“死”，“痊愈”或“未痊愈”，“某公司的品牌”或“其他品牌”，两种结果发生的可能性不必
6 相等.

在本章和下一章，我们将讨论掷硬币、掷骰子、从罐中取筹码、将球放入箱中等几种有价值的模型，其价值在于它们是大量复杂模型的简单而实用的原型，这些复杂模型来自电子物理、心理学、社会学、教育学、生物学、经济学、化学等许多不同领域的实验. Feller(1968)很好地研究了这些模型的差异性. 这些内容本章先做部分介绍，因为需要其他非参数方法的知识，所以大部分说明要放到后面的几个章节中.

事件

这样，我们把掷硬币作为一个试验，并把每一次投掷作为一个基本试验，把一个或多个基本试验或整个试验的结果称为事件（event）. 刚描述的掷硬币试验包含 n 个基本试验，其中每个基本试验的结果是事件“H”或“T”. 事件的组合本身也是一个事件，所以试验中 2^n 个可能结果的每一个都可以看作一个事件. 其他事件的例子有“至少一次正面朝上”、“第4次反面朝上”、“正面朝上的次数至少是反面朝上的两倍”等等.

推而广之，可得如下准则：

准则1.1.1 如果一个试验包含 n 个基本试验，且每一个基本试验有 k 种可能的结果，则整个试验有 k^n 个可能的结果.

例1.1.1

如果一个试验由7个基本试验组成，每个基本试验是向3个盒子之一投一个球. 第一次投会产生3个不同的结果，前两次投共有 $3^2=9$ 种可能的结果. 这样推广到由7次投掷组成的试验，它的试验结果就有 $3^7=2187$ 种不同的可能. ■

现在，考虑一个盒子装有 n 个筹码，筹码编号从1到 n. 从盒子中先取出一个筹码放到桌子上，可以看到它的编号. 这个筹码是从 n 个筹码中任取的一个，所以有 n

种方式. 再从盒子中剩余的筹码中选取第 2 个筹码，并置于第 1 个筹码的旁边，它的编号也能够看到. 由于此前盒子中只剩下 $n-1$ 个筹码，所以第 2 个筹码的选择有 $n-1$ 种方式. 因为选择第 1 个筹码的 n 种方式中的每一种都对应着第 2 个筹码的 $n-1$ 种取法，所以第 1、2 个筹码的选择共有 $n(n-1)$ 种方式. 第 3 个筹码有 $n-2$ 种选取 7
方式，取出后置于桌子上的第 2 个筹码旁边. 现在，依次选取 3 个筹码共有 $n(n-1)(n-2)$ 种方式. 如此继续下去，直至最后一个筹码被取出（最后一个筹码的选取只有一种方式，因为此时盒子中只有 1 个筹码），可知从盒子中选取有 n 个编号的筹码的方法有

$$n(n-1)(n-2)\cdots(3)(2)(1)=n! \tag{1}$$

（读作“n 的阶乘”）种方式，或者说把 n 个不同的物体排成一行的排列方式有 $n!$ 种. 为方便起见，我们定义 $0!=1$.

准则 1.1.2　n 个不同的物体排成一行，其排列方式有 $n!$ 种.

例 1.1.2

考虑字母 A、B、C 排成一行的所有排列数. 第 1 个字母是这 3 个字母中的任意一个，当第 1 个字母选定之后，第 2 个字母有两种不同的选取方式，剩下的字母就是最后一个选取的，这样一共有 $(3)(2)(1)=6$ 种不同的排列. 这 6 种排列是 ABC、ACB、BAC、BCA、CAB 和 CBA. ■

例 1.1.3

假设一个有 8 匹马赛马比赛. 如果你能正确地预测出哪匹马获得第一，哪匹马获得第二，压上对这个结果的赌注，那么你就“赢得正序连赢”. 假设你要确保能赢得正序连赢，这意味着你要买 $(8)(7)=56$ 张彩票，它们分别对应着产生第一、二名的 56 种可能的结果. 整个比赛的结果，即 8 匹马的名次排列，共有 $8!=40\,320$ 种不同的方式. ■

如果 n 个物体互不相同，那么这 $n!$ 种排列中的每一种都是唯一的. 但如果有 2 个物体相同，那么对于这 n 个物体的每一种排列，存在另一种排列的结果与之一致，在这两种排列中，其余 $n-2$ 个物体的位置是一样的，只有两个相同物体的位置交换了. 所以 $n!$ 种排列中的每一个排列都有另一个排列与它相同. 所以不同的排列数是 $n!/2$ 或 $n!/2!$.

假设其中有 3 个物体相同，其余 $n-3$ 个互不相同. 如果我们把这 $n!$ 种排列按相同的排列分组，则在每组中有 $3!$ 个排列. 因为由准则 1.1.2，三个相同的物体在它们的三个位置上有 $3!$ 种不同的且无法区分的排列方式，那么不同排列的数目，等于 8
按相同排列分组的组数，是 $n!/3!$. 如果有 n_1 个物体相同，这 $n!$ 种排列按相同排列分组，每组的容量是 $n_1!$. 如果有 n_1 个相同的类型 1 物体，n_2 个相同的类型 2 物体，对于类型 1 物体的每一种排列，类型 2 有 $n_2!$ 种相同的排列，这样每组中排列方式相同的排列就有 $n_1!\,n_2!$ 种. 因此，分得的组数是 $n!/(n_1!\,n_2!)$. 这就引出如下计数准则：

准则 1.1.3 如果 n 个物体中有 n_1 个相同的第 1 类物体，n_2 个相同的第 2 类物体，……，n_r 个相同的第 r 类物体，则将其排成一行并可以区分的排列数，记作 $\begin{bmatrix} n \\ n_i \end{bmatrix}$，是

$$\begin{bmatrix} n \\ n_i \end{bmatrix} = \frac{n!}{n_1!n_2!\ldots n_r!} \tag{2}$$

特别地，如果 n 个物体由 k 个相同的某类物体和 $n-k$ 个相同的另一类物体组成，那么这 n 个物体排成一行并可区分的排列数，记作 $\binom{n}{k}$，为

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{3}$$

本书约定，如果 k 大于 n，则 $\binom{n}{k}=0$，显然，不存在这 n 个物体的排列使得其中有多于 n 个的物体是一样的.

为了说明准则 1.1.3 的应用，我们把这 $n!$ 种排列按相同的排列分组，每组中有 $n_1!n_2!\cdots n_r!$ 种排法. 因为任何一个排列不会出现在两个不同的组中，所以组数是 $n!/(n_1!n_2!\cdots n_r!)$. 不失一般性，可设 $n_1+n_2+\cdots+n_r=n$（n_i 可以是 1，表示只与自己相像的物体），因为 $1!=1$，方程（2）除以 1 并不影响其数值，且准则 1.1.3 仍然成立. 可以看出准则 1.1.2 是准则 1.1.3 的特例，其中所有 $n_i=1$.

例 1.1.4

例 1.1.2 计算了将字母 A、B、C 排成一行的排列有 6 种方式，现假设字母 A 与 B 相
9 同，用字母 X 表示，则排列 ABC 和 BAC 记作 XXC，是不可区分的. 同样 ACB 和 BCA 记作 XCX，则最初的 $3!=6$ 种排列减为

$$\binom{3}{2} = \frac{3!}{2!1!} = \frac{(3)(2)(1)}{(2)(1)(1)} = 3$$

种可区分的排列，即 XXC、XCX 和 CXX. ■

例 1.1.5

将一枚硬币投掷 5 次，假定结果是两次正面三次反面. 它们出现的不同顺序相当于排列两类物体，一类含 2 个相同物体，另一类含 3 个相同物体，即 $\binom{5}{2}=10$. 这 10 种排列方式如下，其中 H 表示正面，T 表示反面.

HHTTT　THHTT　TTHHT　HTHTT　THTHT

TTHTH　HTTHT　THTTH　TTTHH　HTTTH ■

从 n 个物体中选取 k 个物体可以形成多少组呢？我们用准则 1.1.3 来回答这个问题. 假设 n 个物体排成一排，将 k 个相同的标记放在这 n 个物体中的 k 个上，容易看出，

这样的放法数，等于 k 个标记的位置和 $n-k$ 个未标记的位置的可辨别的排列数，是 $\binom{n}{k}$，这正是准则 1.1.3 所给出的．在这种情况下，$\binom{n}{k}$ 常被读作“从 n 个物体中一次取 k 个的取法数”．

例 1.1.6

考虑字母 A,B,C，从中选取两个字母的取法有 $\binom{3}{2}=3$ 种，即 AB、AC 和 BC．为了与前面的讨论联系起来，把三个字母中的两个用“*”标记：

A* B* C 得到 AB

A* B C* 得到 AC

A B* C* 得到 BC ■

注意它与例 1.1.4 的相似性． 10

二项式系数

我们介绍 $\binom{n}{k}$ 的另一种使用方法，考虑表达式 $(x+y)^n=(x+y)(x+y)\cdots(x+y)$，$x^n$ 项是由第一个因式中的 x 项，第二个因式中的 x 项……直至第 n 个因式中的 x 项相乘得到的．$x^{n-1}y$ 项是由 $n-1$ 个因式中的 x 项相乘，再乘以一个因式中的 y 项得到的，因为 y 项可以从这 n 个因式中的任一个选取，所以 $(x+y)^n$ 展开式中有 n 项 $x^{n-1}y$．同理，对于每一个 k 值，x^ky^{n-k} 项是从 k 个因式中选取 k 个 x 项，再从其余的 $n-k$ 个因式中选取 $n-k$ 个 y 项得到的．为得到 x 项而选取 k 个因式的方法有 $\binom{n}{k}$ 种，其余的因式得到的是 y 项．这样在 $(x+y)^n$ 的展开式中 x^ky^{n-k} 项出现了 $\binom{n}{k}$ 次，将所有的展开项加起来得到：

$$(x+y)^n=x^n+\binom{n}{n-1}x^{n-1}y^1+\binom{n}{n-2}x^{n-2}y^2+\ldots$$
$$+\binom{n}{2}x^2y^{n-2}+\binom{n}{1}x^1y^{n-1}+y^n \tag{4}$$

注意 $0!=1$，所以 $\binom{n}{0}=1$ 且 $\binom{n}{n}=1$．如果使用记号

$$\sum_{i=a}^{b}C_i=C_a+C_{a+1}+C_{a+2}+\cdots+C_{b-1}+C_b$$

读作“i 从 a 到 b，对 C_i 求和”，(4) 式可记为

$$(x+y)^n=\sum_{i=0}^{n}\binom{n}{i}x^iy^{n-i} \tag{5}$$

这就是我们熟知的“二项式展开（binomial expansion）”，在很多高中数学的教科书上都可以找到它. 这也解释了为什么“二项式系数（binomial coefficient）”常用来描
11 述记号$\binom{n}{k}$. 同理，在（$x_1+x_2+\cdots+x_r$）^{n} 的展开式中 $x_1^{n_1}x_2^{n_2}\cdots x_r^{n_r}$ 的系数由“多项系数（multinomial coefficient）” $\left[\begin{matrix}n\\n_i\end{matrix}\right]$给出.

例 1.1.7

利用二项式展开来计算（2+3）^{4}，当然，我们知道答案是 $5^4=625$. 由（5）式中的二项式展开有

$$\begin{aligned}(2+3)^4&=\sum_{i=0}^{4}\binom{4}{i}2^i3^{4-i}\\&=\binom{4}{0}2^03^4+\binom{4}{1}2^13^3+\binom{4}{2}2^23^2+\binom{4}{3}2^33^1+\binom{4}{4}2^43^0\\&=(1)(1)(81)+(4)(2)(27)+(6)(4)(9)+(4)(8)(3)+(1)(16)(1)\\&=81+216+216+96+16=625\end{aligned}$$

■

习题

1. 用 0 至 9 这 10 个数字可以组成多少个 4 位数（从 0000 到 9999）？其中每个数字可以重复多次.
2. 利用字母表中的 26 个字母，可以得到多少种 4 个字母的排列？其中每个字母可以重复使用多次.
3. 将字母 L,O,V,E 排成有四个字母的“单词”，每个字母只能用一次，共有多少种排法？
4. 5 个人坐成一排有多少种坐法？
5. 从一个 12 人的俱乐部中选取 3 人组成一个委员会，有多少种选法？
6. 在$(x+y)^6$ 的展开式中，x^3y^3 项的系数是多少？
7. 在$(x+y+z)^7$ 的展开式中，x^2y^4z 项的系数是多少？
8. 在$(w+x+y+z)^7$ 的展开式中，x^2y^5 项的系数是多少？（提示：$x^2y^5=w^0x^2y^5z^0$）
9. 计算 $\sum_{i=1}^{3}\binom{4}{i}$.
10. 计算 $\sum_{i=0}^{3}\binom{4}{i}\left(\frac{1}{2}\right)^2$.
11. 计算 $\sum_{i=3}^{5}\binom{6}{i}\left(\frac{1}{3}\right)^i\left(\frac{2}{3}\right)^{6-i}$.

12 12. 计算 $\sum_{i=1}^{4}5$.

思考题

1. 从第一类中选取 n_1 个物体，从第二类中选取 n_2 个物体，依次进行，从第 r 类中选取 n_r 个物体，共有多少种选择方式？其中第一类物体共有 N_1 个，第二类物体共有 N_2 个，等等. 如果对于某个 i, n_i 比 N_i 大时，又有多少情况呢？
2. 证明 $\sum_{i=0}^{n}\binom{n}{i}p^i(1-p)^{n-i}=1$.

1.2 概　率

本节将应用 1.1 节中的 3 个计数准则去发现一些有趣而有用的概率. 首先介绍一些统计学中的标准术语，正确理解本节和其他节定义中的术语，能使我们更容易理解统计概念.

样本空间

我们从试验出发来定义重要的术语：*样本空间*（sample space）和*样本空间中的点*（a point in the sample space）.

定义 1.2.1　*样本空间*是一个试验中所有可能的不同结果的集合.

定义 1.2.2　*样本空间中的一个点*是一个试验中一次可能的结果.

每个试验都有其自己的样本空间，它含有这个试验所有可能的不同结果，通常假定样本空间可能有一个足够细密且合理的划分，其中每一个划分称为一个点，且假定每一个可能的结果用且仅用一个点来表示.

例 1.2.1

如果一个试验为掷两次硬币，则样本空间中有 4 个点：HH、HT、TH 和 TT. ■

例 1.2.2

以对一个学生进行 10 个判断题的测试作为一个试验，每个判断题答案为“对或错”，则样本空间中有 $2^{10}=1024$ 个点，每一个点是对这 10 个连续问题可能答案的序列，比如“TTFTFFTTTT”. ■ 13

事件

有了样本空间中的点，我们可以定义*事件*（event）.

定义 1.2.3　一个*事件*是样本空间中一些点的集合.

在例 1.2.1 中我们提过“两次正面”这个事件，它是由 HH 这一个点组成的；事件“一次正面”含有 TH 和 HT 两个点；事件“至少一次反面”含有 TH，HT 和 TT 三个点；同样，事件“四次正面”中没有点. 通常称一个没有点的集合为*空集*（empty set）. 一个包含样本空间中所有点的事件称为*必然事件*（sure event），因为每

次试验时该事件必然会发生.

两个不同的事件可能含有相同的点，事件“至少一次反面”和事件“至少一次正面”就有两个相同的点 TH 和 HT. 如果两个事件没有相同的点，则称它们为*互不相容*（mutually exclusive）事件，因为其中的一个事件的发生排除了另一个事件同时发生的可能.

如果一个事件中的所有点都包含在另一个事件中，那么称第一个事件*包含于*（contained in）第二个事件，或者称第二个事件包含第一个事件. 事件“至少一次正面”包含事件“两次正面”. 每个事件当然包含它自身.

概率

样本空间中的每一个点对应于它的一个数，这个数称为*点的概率*（probability of the point）或*结果的概率*（probability of the outcome）. 概率是 0 至 1 中的任意一个实数. 如果我们在相同的条件下多次重复该试验，那么这个点或事件发生的频数（frequency）就是这个点或事件发生概率的一个近似值.

定义 1.2.4　如果 A 表示一个试验中的事件，n_A 是这个试验独立重复了 n 次中事件 A 发生的次数，那么*事件 A 发生的概率*，记作 $P(A)$，由下式给出

$$P(A)=\lim_{n\to\infty}\frac{n_A}{n} \tag{1}$$

14 读作“当重复试验次数趋于无穷时，事件 A 发生的次数与试验重复的次数之比的极限”.

独立（independent）的正式定义将在后面给出. 现在我们暂且认为，如果某个试验的结果不影响其他试验的结果，则这些试验是相互独立的.

事件发生概率的定义包含了单点发生概率的定义，后者可作为前者的一个特例，因为一个事件也可以由单点构成. 由于事件发生的次数等于组成该事件所有互不相容结果发生的次数之和，从定义可知，显然一个事件发生的概率等于组成该事件所有结果发生的概率之和.

概率函数

在实际应用中，一个特定样本空间的概率集合是很少已知的，但根据试验者的先验知识，这些概率应是确定的，即试验者建立一个模型来作为这个试验的理想化描述，然后察看这个试验模型的样本空间，并用某种合理的方式来给样本空间中不同点的赋概.

例 1.2.3

在掷一次均匀硬币的试验中，一般认为结果 H 发生一半次数的假设是合理的. 这样可给结果 H 赋概 1/2，对结果 T 也一样. 我们记作 $P(\mathrm{H})=1/2, P(\mathrm{T})=1/2$. ■

例 1.2.4

在掷三次均匀硬币的试验中，假设 $2^3=8$ 个结果 HHH、HHT、HTH、HTT、THH、THT、

TTH、TTT 中的每一个发生具有等可能性是合理的，即每一个结果发生的概率是 1/8，则可得到 P(三次反面) = 1/8, P(至少一次正面) = 7/8，P(正面多于反面) = P(至少两次正面) = 4/8 = 1/2. ■

前两个例子中我们已经用到了**概率函数**（probability function）.

定义 1.2.5 概率函数是指对样本空间中不同事件指定概率的函数.

在例 1.2.3 中，概率函数是由 $P(\mathrm{H}) = 1/2, P(\mathrm{T}) = 1/2$ 给出. 概率函数必须对样本空间中的每个点都给出概率，而且样本空间中事件的概率可由它所包含样本点的概率来确定. 15

概率函数有些性质很明显. 设 S 为一样本空间，A 为 S 中的事件，如果 P 是一个概率函数，由

$$P(S) = \lim_{n\to\infty}\frac{n}{n} = 1$$

得 $P(S) = 1$. 因为 $n_A \geqslant 0$，所以

$$\lim_{n\to\infty}\frac{n_A}{n} \geqslant 0$$

得到 $P(A) \geqslant 0$. 记 $\bar{A}$ 为事件“A 不发生”，因为 $n_{\bar{A}} = n - n_A$，并且

$$\lim_{n\to\infty}\frac{n_{\bar{A}}}{n} = \lim_{n\to\infty}\frac{n-n_A}{n} = \lim_{n\to\infty}\left(1-\frac{n_A}{n}\right) = 1 - \lim_{n\to\infty}\frac{n_A}{n} = 1 - P(A)$$

所以得 $P(\bar{A}) = 1 - P(A)$,

条件概率

前面提过，一个试验各种不同的结果是互不相容的，然而与一个试验关联的不同事件却可能没有这个性质. 在掷三次硬币的试验中，事件“三次正面”与事件“至少两次正面”很可能同时发生. 考虑在事件“至少两次正面”发生的条件下，事件“三次正面”发生的概率. 如果至少两次正面朝上发生了，则样本空间中的一些点，比如，TTT, TTH, THT 和 HTT 可以排除，试验的可能结果就减少为 4 个等可能的点. 所以，每一个点的概率变成 1/4，因此，在给定至少有两次正面朝上的事件发生时，事件“三次正面”或 HHH 的概率是 1/4. 已知的附加信息起到了排除掉一些结果的作用，也就人为减小了样本空间.

考虑另一个掷骰子试验. 设 S 为样本空间，A 是“4,5 或 6 发生”的事件，B 是“偶数（2,4 或 6）发生”的事件，如图 1-1 所示. 在事件 B 发生的条件下，事件 A 16
发生的概率记作 $P(A|B)$，通常读作“给定事件 B 发生条件下事件 A 发生的概率”. 既然已知 B 发生，我们不仅可以去掉既不在 A 中又不在 B 中的点，即点 1 和 3，而且还可去掉不在 B 中但在 A 中的点，即点 5. 这样全部去掉了不在 B 中的点，此时样本空间就是 B 中点的集合. B 中能使事件 A 也发生的点是既在 A 中又在 B 中的点，为点 4 和点 6，它表示“事件 A 和 B 同时发生”.

定义 1.2.6 如果 A, B 是样本空间 S 中的两个事件，事件“A, B 同时发生”由

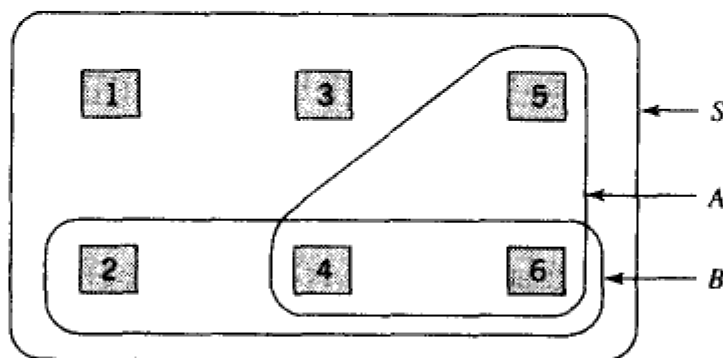


图1-1 样本空间 S 中的事件 A，事件 B

样本空间中同时在 A 和 B 中的点组成，称事件 A 与事件 B 的交（joint event A and B），记作 AB. 事件交（joint event）的概率记作 $P(AB)$.

“给定 B 发生条件下 A 发生”的概率是“AB”发生的概率除以样本空间“B”的概率，或用符号表示为：

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{2}$$

从另一方面来看，假设前面的试验进行了 n 次，且仅记录下了事件 B 中的结果，而不发生在事件 B 中的结果没有记录. 设 n_B 为事件 B 发生的次数，n_{AB} 为事件 B 发生时 A 发生的次数，那么

$$P(A|B)=\lim_{n\to\infty}\frac{n_{AB}}{n_B}=\lim_{n\to\infty}\frac{n_{AB}/n}{n_B/n}=\frac{P(AB)}{P(B)} \tag{3}$$

如此我们从直观上证明了下面的定义：

定义 1.2.7 给定事件 B,A 发生的条件概率（conditional probability）就是当给定事件 B 发生时事件 A 发生的概率，由下式给出：

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{4}$$

其中 $P(B)>0$. 如果 $P(B)=0$，则 $P(A|B)$ 无定义.

例 1.2.5

考虑投掷一颗匀称的骰子，6 种可能结果中每一种发生的概率都是 1/6. 设 A 是“4,5 或
17 6 发生”的事件，B 是“偶数发生”的事件. 那么，$P(AB)=P(4\text{ 或 }6)=2/6=1/3$，并且 $P(B)=3/6=1/2$，则条件概率 $P(A|B)$ 由下式算出：

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{1/3}{1/2}=\frac{2}{3}$$ ■

我们应该注意这个答案的合理性，由于知道偶数（事件 B）已经发生了，则试验可能结果是 2,4 或 6. 我们想知道一个比 3 大的数（事件 A）发生的可能性，由于这三个偶数中有两个比 3 大，故答案是 2/3.

独立事件

条件概率的思想很自然地引出独立事件（independent event）的思想出现. 如果在事件 B 发生的条件下，A 发生的概率与在不知道 B 是否发生的情况下 A 发生的概

率相等，我们认为事件 A 的发生与否与事件 B 的发生与否独立，即如果 $P(A|B)=P(A)$，则 A 独立于 B. 事实上，这就是事件独立性的常用定义，但这种定义从形式上看不出 B 是否独立于 A. 所以我们在条件概率（4）式中用 $P(A)$ 代替 $P(A|B)$，得到如下定义：

定义 1.2.8 称两个事件 A、B 是*独立的*（independent），如果

$$P(AB)=P(A)P(B) \tag{5}$$

由（5）式的对称性，容易看出，如果 A 独立于 B，那么 B 也独立于 A，这时我们说"A 与 B 是独立的"，即是说它们相互独立.

例 1.2.6

在投掷两次均匀硬币的试验中，样本空间中的 4 个点的发生是等概率的. 设事件 A 是"第一次正面朝上"，事件 B 是"第二次正面朝上"，则 A 中有点 HH 和 HT，B 中有点 HH 和 TH，且 AB 中有点 HH. 所以 $P(A)=2/4$，$P(B)=2/4$，且 $P(AB)=1/4$，满足（5）式，故 A 和 B 独立. ■

下面的例子说明两个事件的独立性并不总是直观的，需要直接由定义和（5）式来验证. 18

例 1.2.7

仍然考虑投掷一个均匀骰子的试验，样本空间中含有 6 个等可能的点 1,2,3,4,5 和 6. 设 A 是事件"偶数发生"，包含点 2,4 和 6；B 是事件"至少为 4 的数"，包含点 4,5 和 6；C 是事件"至少为 5 的数"，包含点 5 和 6. 因为 $P(A)P(B)=(1/2)(1/2)=1/4$，而 $P(AB)=1/3$，所以 A 和 B 是不独立的. 然而 A 和 C 是独立的，因为 $P(A)P(C)=(1/2)(1/3)=1/6$，与 $P(AC)$ 相等. ■

*独立事件*和*互不相容事件*的概念有时可能会产生混淆，因为两个概念都给人一种"两个事件互不干扰"的印象. 独立事件的概念不仅依赖于所考虑的两个事件，而且还依赖于定义于同一概率空间上的概率函数，对于某个概率集合有可能 $P(AB)$ 与 $P(A)P(B)$ 相等，而对于另一概率集合有可能 $P(AB)$ 与 $P(A)P(B)$ 不相等. 但是"互不相容"是简单地指两个事件没有共同点，而不管定义在同一空间上的概率函数如何，即 AB 是空集，所以 $P(AB)=0$. 如果 A 和 B 是互不相容的，由（5）式可知，只有在 $P(A)$ 或 $P(B)$ 为零的情况下，事件 A，B 才是相互独立的.

独立试验

我们给出*独立试验*（independent experiment）的如下定义.

定义 1.2.9 两个试验称作是独立的，如果一个试验中的任一事件 A 与另一个试验中的任一事件 B 都满足下式：

$$P(AB)=P(A)P(B)$$

两个试验独立的定义等价于一个试验中的每一事件都独立于另一个试验中的每一事件.

验证对应于两个试验中的每一对事件是否满足定义 1.2.9 是比较繁琐的. 然而，按定义来验证那些仅含有一个点的事件就足够了，对于其他事件的情形也自然得到了验证.

实际应用中，模型的建立通常假设了独立性，然后再利用独立性及定义 1.2.9 中
19 的 $P(A)$ 和 $P(B)$ 来计算 $P(AB)$，这也是独立性定义的主要价值. 因此，独立试验的定义可合理地推广到含有多于两个试验的情形.

定义 1.2.10　称 n 个试验相互独立（n experiments are mutually independent），如果从这 n 个试验中的每个试验中任取一个事件组成 n 个事件的集合，都满足如下等式：

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n) \tag{6}$$

其中 A_i 表示第 i 个试验中的某个结果，$i=1,2,\cdots,n$.

如果不引起混淆的话，可从前面定义中省略“相互”这个词.

例 1.2.8

考虑投掷一次不均匀硬币的试验，其中事件 H 发生的概率是 p，事件 T 发生的概率是 $q=1-p$. 独立重复三次这个试验，可用下标来记录相关试验的结果，其中 $H_1T_2H_3$ 表示第 1 次试验结果是 H，第 2 次是 T，第 3 是 H. 由于独立性假设，有

$$P(H_1T_2H_3)=P(H_1)P(T_2)P(H_3)=pqp$$

如果我们考虑这 3 次试验中事件“恰有两次出现正面”，则有 $\binom{3}{2}=3$ 种方式，因此

$$P(\text{恰有两次出现正面})=3p^2q$$ ■

很显然，上面的试验可以用 3 个独立基本试验来描述，同时也可推广到一个含有 n 个独立投掷的试验，得到“恰有 k 次出现正面”的概率等于 p^kq^{n-k} 乘以该项所出现的次数. 因此，对于 n 次独立投掷一个硬币试验，有

$$P(\text{恰有}k\text{次出现正面})=\binom{n}{k}p^kq^{n-k} \tag{7}$$

其中，对于每一次投掷 $p=P(H)$.

前面的 3 种定义可以从正反两个方面加以说明，对所有的定义也是如此. 例 1.2.6 给出了（5）式满足时，两个事件是独立的. 例 1.2.8 给出了在有独立性假设时，（6）式满
20 足. 这就是说，如果（6）式不满足，那么这些试验就不是独立的；反之，如果这些试验不独立，则至少有一事件的集合 $A_1A_2\cdots A_n$ 不满足（6）式.

习题

1. 在投掷 3 次硬币的试验中，考虑投掷的次序（从第 1 次至第 3 次），列出样本空间中的点.
2. 根据习题 1，给出
   (a) 两个互不相容事件.

(b) 两个非互不相容的事件.

3. 如果降雨概率是0.15，那么不降雨的概率是多少?
4. 如果到达一个十字路口时，绿灯亮的概率是0.35，黄灯亮的概率是0.10，那么红灯亮的概率是多少?
5. 如果一个足球队赢球与输球的概率是相等的（假设没有平局发生，并且每场比赛的结果是独立的)，那么在整个有8场比赛的赛季中，这个球队至少输掉7场比赛的概率是多少?
6. 如果一个球队赢得每场比赛的概率是0.4，且与其他场次比赛独立，假设整个赛季有10场比赛，那么这个球队至多赢一场的概率是多少?
7. 如果得到一张破损美元纸币的概率是0.05，那么在得到的三张纸币中有两张是破损的概率有多大（假设满足独立性)?
8. 如果被盗的汽车中有60%能找回，且每年有2%的汽车被盗，那么一个人的汽车被盗且再也找不回来的概率是多大?
9. 顾客购买某种品牌清洁器的概率是0.15，有40%的顾客在购买清洁器时会购买一个散雾器，那么一个顾客同时购买这两种商品的概率是多少?
10. 在投掷一枚均匀硬币的3次独立试验中，至少有一次反面朝上的概率是多大?
11. 在投掷一枚均匀硬币的3次独立试验中，若已知至少有一次已经正面朝上，那么这时3次正面朝上的概率是多大?
12. 在投掷一枚均匀硬币的4次独立试验中，若已知至少有2次已经正面朝上，那么这时至少3次正面朝上的概率是多大?
13. 在投掷一枚均匀硬币的3次独立试验中，如果已知第1次得到的是正面，那么这时3次正面朝上的概率是多大?（注：习题11和13的答案不同.）
14. 若每年银行的学生账号中有75%被关闭，且银行20%的账号是学生账号，那么银行的一个账号是学生账号并且使用期超过一年的概率是多大? 21
15. 在投掷一枚均匀硬币的4次独立试验中，若已知至少1次已经反面朝上，那么这时得到至少3次正面朝上的概率是多大?
16. 一个摸彩游戏是从0到9中随机（有放回）地选取3个数字，3次抽取相互独立，例如，212或935.
    (a) 在一次尝试中，成功猜对3个数字的概率是多大? 3个数字不考虑次序，例如，即使抽出数的次序是512或152，那么对215的猜测也算对.
    (b) 猜对555的概率与猜对212或935的概率一样吗?

思考题

1. 证明：在一个有n个点的样本空间中，至少含有一个点的事件的数目是2^n-1.
2. 本题意在表明两两独立并不意味着三个整体独立. 考虑从装有数字1至9的帽子中取数，等可能取到每个数字的可能性相等. 如果取到1,2或3，则事件A发生；取到1,4或5，则事件B发生，取到2,4或6，则事件C发生.
    (a) 证明事件A与B独立；A与C独立；以及B与C独立.
    (b) 事件A,B,C相互独立吗? 为什么不是?

（c）求事件 A,B,C 都不发生的概率.

1.3 随机变量

随机变量

一个试验的结果可能是数值的，比如考试的分数；也可能是非数值的，如从围栏中逃出的老鼠对“红色的门”的选择. 为了分析试验的结果，常用的做法是对样本空间中的点赋值. 任何一个这样赋值的准则称为随机变量（random variable）.

22

定义 1.3.1 随机变量是定义在样本空间上点的实值函数.

通常用大写字母 W,X,Y 或 Z 来表示随机变量，也可以带下标. 随机变量的值用小写字母表示.

例 1.3.1

在一个试验中，顾客可以从3种商品即肥皂、清洁剂或商标A中选取一种，样本空间包含代表3种可能选择的3个点. 若选择商标A，则随机变量 X 取值1；若是其他两种结果则取值为0. 所以 $P(X=1)$ 为顾客选择商标A的概率. ■

在一个样本空间定义多个随机变量是比较方便的做法，如下面例题所示.

例 1.3.2

调查6个女孩和8个男孩，看哪一个更容易与自己的母亲还是父亲交流. 设 X 表示与母亲更容易交流的女孩数，Y 表示与母亲更容易交流的孩子总数. 如果 $X=3$，则事件“3个女孩觉得与母亲交流更容易”发生，若与此同时，$Y=7$，则事件“3个女孩和 $7-3=4$ 个男孩觉得与母亲交流更容易”发生. ■

若 X 是一个随机变量，“$X=x$”表示样本空间中相应事件的简化符号，它是包含随机变量 X 取值为 x 的所有点的集合.

例 1.3.3

掷一枚硬币两次，设 X 为正面朝上的次数，那么“$X=1$”对应着仅含有点HT和TH的事件. ■

因此“$X=x$”通常称作“事件 $X=x$”，意思是“随机变量 X 取值为 x 的所有结果的事件”.

由于随机变量与事件之间的这种紧密关系，条件概率（conditional probability）与独立性（independence）的定义同样可用于随机变量.

23

定义 1.3.2 给定 Y 时 X 的条件概率是指在随机变量 Y 取值为 y 时，随机变量 X 取值为 x 的概率，记为 $P(X=x|Y=y)$.

由定义 1.2.7 可知，条件概率的公式可表示如下：

$$P(X=x|Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)} \quad 若\ P(Y=y)>0 \tag{1}$$

例 1.3.4

在例 1.3.2 中，设 X 为 6 个女孩中觉得与母亲交流更容易的女孩数，Y 为更容易与母亲交流的孩子总数. 为方便起见，设 $Z=Y-X$，它是 8 个男孩中觉得更易与母亲交流的男孩数目. 假设每个孩子的答案是相互独立的，并且每个孩子说他（或她）更易与其母亲交流的概率是 p（未知）. 求条件概率 $P(X=3\mid Y=7)$.

首先由假设可知，$X=3$ 和 $Z=4$ 是独立的. 因为事件（$X=3,Y=7$）和事件（$X=3,Z=4$）是一样的，所以联合概率为（由例 1.2.8）：

$$\begin{aligned}P(X=3,Y=7)&=P(X=3,Z=4)\\&=P(X=3)P(Z=4)\quad \text{由独立性}\\&=\binom{6}{3}p^3(1-p)^3\binom{8}{4}p^4(1-p)^4\end{aligned}\tag{2}$$

同样，我们可得到

$$P(Y=7)=\binom{14}{7}p^7(1-p)^7\tag{3}$$

所以条件概率 $P(X=3\mid Y=7)$ 为

$$P(X=3|Y=7)=\frac{P(X=3,Y=7)}{P(Y=7)}=\frac{\binom{6}{3}\binom{8}{4}}{\binom{14}{7}}=.408\tag{4}$$

因为有关未知数 p 的因式彼此抵消掉了. ■ 24

概率函数

正如一个样本空间中的点是互不相容的，随机变量的取值也是互不相容的，即对于试验的一个结果，所定义的随机变量也只有一个取值. 这样随机变量所有取值的集合与样本空间有许多相同的性质. 随机变量各个单独的取值对应于样本空间中的各个点，一个取值集合则对应着一个事件，随机变量在一个数值集合内取值的概率等于它在这个集合内取每一个值的概率之和，例如：

$$P(a<X<b)=\sum_{a<x<b}P(X=x)$$

式中求和项包括所有 a、b 之间的 x，但不包含 a、b，且

$$P(X=\text{偶数})=\sum_{x\text{为偶数}}P(X=x)$$

其中 $\sum$ 是对所有取值为偶数的 x 求和. 由于 X 的取值集合与样本空间的相似性，描述 X 各种可能取值的概率称为“随机变量 X 的概率函数（probability function of the random variable X）”，正如一个样本空间有一个概率函数一样. 但随机变量的概率函数不是概率的任意赋值，如样本空间的概率函数，因为样本空间的每个点一旦赋予了概率且在样本空间上定义了随机变量，那么就知道了 X 各种取值的概率，从而 X 的概率函数也相应确定.

定义 1.3.3 对任一实数 x，随机变量 X 的概率函数就是 X 取值为 x 的概率，常用 $f(x)$ 表示. 换句话说,

$$f(x)=P(X=x) \tag{5}$$

在 X 不能取到的那些 x 处概率函数为 0.

有时候，用条形图来表达概率函数是很方便的，即用随机变量的取值作为横坐
25 标（沿着水平轴），以概率为纵坐标（条形的高度）. 例如，如果 $P(X=1)=0.3$, $P(X=2)=0.4$, $P(X=4)=0.3$，则这个概率函数的条形图如图 1-2 所示，条形的高度表示随机变量取各种值的概率

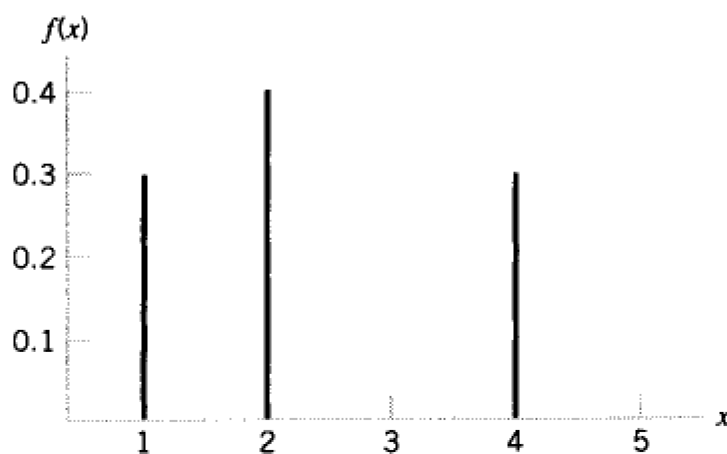


图 1-2 条形图

$f(x)$ 只是概率函数的一种表示方式，其他的表示有 $f_0(x)$, $f_1(x)$, $f_2(x)$, $g(x)$, $h(x)$ 等等. 但不同表达式的含义应从上下文了解.

分布函数

我们已经看到，随机变量的概率分布可以用一个概率函数来描述，还有另一种描述方法，就是分布函数（distribution function），它描述累积概率.

定义 1.3.4 对任意的实数 x，随机变量 X 的分布函数就是 X 取值不大于 x 的概率，通常记作 $F(x)$，换句话说,

$$F(x)=P(X\leqslant x)=\sum_{t\leqslant x}f(t) \tag{6}$$

其中求和是对所有不超过 x 的 t 进行的. 分布函数也称作累积分布函数（简记 c. d. f）以强调它表示累积概率.

26 分布函数也可以用图形来表示，x 作为横坐标，$F(x)$ 作为纵坐标. 沿用前面的例子，假设 $P(X=1)=0.3$, $P(X=2)=0.4$, $P(X=4)=0.3$，则 $F(x)$ 的图形如图 1-3 所示：

图中实际上只含有水平线，画出垂直线只是给这幅图某种直观上的“连接”，还可以帮助找到下一节将要介绍的分位数（quantile）. 垂直线的高度与概率函数条形图中条形的高度相同.

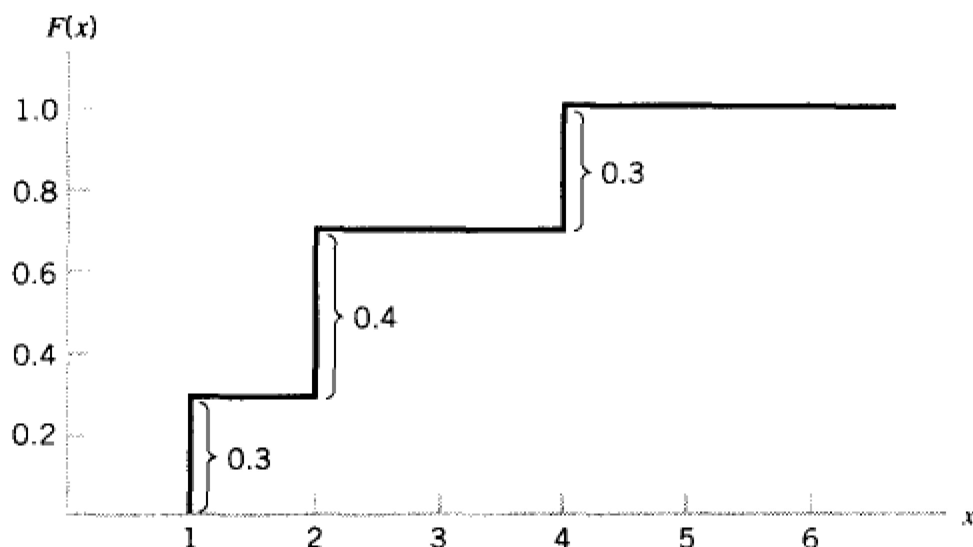


图 1-3 分布函数

二项分布

对一些众所周知的分布，我们给出它们的名字.

定义 1.3.5 设 X 是一个随机变量，二项分布（binomial distribution）是由如下概率函数所确定的概率分布

$$f(x)=P(X=x)=\binom{n}{x}p^xq^{n-x} \qquad x=0,1,\ldots,n \tag{7}$$

其中 n 是正整数，$0\leqslant p\leqslant 1, q=1-p$. 注意我们使用了约定 $0!=1$.

则它的分布函数是

$$F(x)=P(X\leqslant x)=\sum_{i\leqslant x}\binom{n}{i}p^iq^{n-i} \tag{8}$$

其中，求和是对不大于 x 所有可能的 i 进行的. 表 A3（见附表）给出了参数 n 和 p 的某些特定值时 $F(x)$ 的值. 27

例 1.3.5

在含有 n 次独立基本试验的试验中，每个基本试验有两种结果："成功"或"失败"，概率分别是 p 和 q，就像掷硬币一样. 设 X 为在 n 次独立基本试验中"成功"的总次数. 正如 1.2 节中（7）式所示

$$P(X=x)=\binom{n}{x}p^xq^{n-x}$$

其中，x 为 0 到 n 的整数. 所以 X 服从二项分布. ■

离散均匀分布

另一个有用的概率分布就是离散均匀分布（discrete uniform distribution）.

定义 1.3.6 设 X 为一个随机变量，离散均匀分布是由如下概率函数所确定的概率分布：

$$f(x)=\frac{1}{N} \qquad x=1,2,\ldots,N \tag{9}$$

若 X 服从离散均匀分布，则 X 等概率取到 1 至 N 中的每个整数值.

例 1.3.6

罐中装有 N 个塑料筹码，编号从 1 到 N. 试验是每次从罐中取出一个筹码，且取到每个筹码的概率相同. 此样本空间由 N 个点组成，分别表示取出的 N 个筹码. 设 X 为所取出筹码上的编号，则 X 服从离散均匀分布. ■

联合分布

当在同一个样本空间上定义多个随机变量，或对多个试验组成的联合试验并在每一试验上定义了一个或多个随机变量时，需要考虑联合分布，通常用联合概率函数（joint probability function）或联合分布函数（joint distribution function）来刻画.

定义 1.3.7 随机变量 $X_1,X_2,\cdots,X_n$ 的联合概率函数（joint probability function）$f(x_1,x_2,\cdots,x_n)$ 是 $X_1=x_1,X_2=x_2$，$\cdots$，以及 $X_n=x_n$ 同时发生的概率，即如下等式：

28
$$f(x_1,x_2,\ldots,x_n)=P(X_1=x_1,X_2=x_2,\ldots,X_n=x_n) \tag{10}$$

定义 1.3.8 随机变量 $X_1,X_2,\cdots,X_n$ 的联合分布函数 $F(x_1,x_2,\cdots,x_n)$ 是 $X_1\leqslant x_1$，$X_2\leqslant x_2,\cdots$，以及 $X_n\leqslant x_n$ 同时发生的概率，即如下等式：

$$F(x_1,x_2,\ldots,x_n)=P(X_1\leqslant x_1,X_2\leqslant x_2,\ldots,X_n\leqslant x_n) \tag{11}$$

例 1.3.7

考虑例 1.3.2 中定义的随机变量 X 和 Y. 设 $f(x,y)$ 和 $F(x,y)$ 分别是 (X,Y) 的联合概率函数和联合分布函数. 由例 1.3.4 有

$$f(3,7)=P(X=3,Y=7)=\binom{6}{3}\binom{8}{4}p^7(1-p)^7 \tag{12}$$

且

$$F(3,7)=P(X\leqslant 3,Y\leqslant 7)=\sum_{\substack{0\leqslant x\leqslant 3\\ x\leqslant y\leqslant 7}} f(x,y) \tag{13}$$

其中

$$f(x,y)=\binom{6}{x}p^x(1-p)^{6-x}\binom{8}{y-x}p^{y-x}(1-p)^{8-(y-x)}$$

（13）式是对所有满足：$x\leqslant 3$ 且 $y\leqslant 7$，且 $x,y-x$ 都为非负整数这样的 x,y 求和的.

注意 p 值未知，（12）式与（13）式是无法计算的. ■

定义 1.3.9 给定 Y 时，X 的条件概率函数（conditional probability function）$f(x|y)$，是

$$f(x|y)=P(X=x|Y=y) \tag{14}$$

从（1）式我们可得

$$\begin{aligned}f(x|y)=P(X=x|Y=y)&=\frac{P(X=x,Y=y)}{P(Y=y)}\\&=\frac{f(x,y)}{f(y)}\end{aligned} \tag{15}$$

其中，$f(x,y)$是 X 和 Y 的联合概率函数，$f(y)$是 Y 本身的概率函数. 29

例 1.3.8

续例 1.3.7，设$f(x|y)$为给定 $Y=y$ 时 X 的条件概率函数，则由（4）式有

$$f(3|7)=P(X=3|Y=7)=0.408$$

为找到一般情况下（即对任意选取的 x 和 y）$f(x|y)$的表达式，先假设$f(x,y)$表示 X 和 Y 的联合概率函数，则由例 1.3.7 得

$$f(x,y)=\binom{6}{x}p^x(1-p)^{6-x}\binom{8}{y-x}p^{y-x}(1-p)^{8-(y-x)}$$

它最初是（2）的一般形式. 记$f(y)$是 Y 的概率函数，再从例 1.3.4 我们可得：

$$f(y)=P(Y=y)=\binom{14}{y}p^y(1-p)^{14-y}$$

由定义 1.3.9 可得给定 $Y=y$ 时 X 的条件概率函数：

$$f(x|y)=\frac{f(x,y)}{f(y)}=\frac{\binom{6}{x}\binom{8}{y-x}}{\binom{14}{y}}\quad \begin{array}{l}0\leqslant x\leqslant 6\\ 0\leqslant y-x\leqslant 8\end{array} \tag{16}$$

其中轻易地消去了含有未知参数 p 的项. ■

超几何分布

前一个例子中，我们已处理过一个称为*超几何分布*（hypergeometric distribution）的概率函数. 更一般地，通常假设有两类物体，一类有 A 个物体和另一类有 B 个物体（如前面例子中女孩总数和男孩总数），且选中每个物体的可能性相同，则在从 $A+B$个物体中选出 k 个的条件下，有 x 个来自 A 个物体的概率，这就引出了超几何概率函数.

定义 1.3.10 设 X 是一个随机变量，*超几何分布*是由下面的概率函数所确定的概率分布，表示为

$$f(x)=P(X=x)=\frac{\binom{A}{x}\binom{B}{k-x}}{\binom{A+B}{k}}\quad \begin{array}{l}0\leqslant x\leqslant A\\ 0\leqslant k-x\leqslant B\end{array} \tag{17}$$

式中，A,B 和 k 都是非负整数，且 $k\leqslant A+B$. 30

相互独立的随机变量可按照独立试验中的定义 1.2.9 和 1.2.10 的相同的方式来定义.

定义 1.3.11 设随机变量 $X_1,X_2,\cdots,X_n$ 分别具有概率函数 $f_1(x_1),f_2(x_2),\cdots,f_n(x_n)$，且联合概率函数为$f(x_1,x_2,\cdots,x_n)$，那么 $X_1,X_2,\cdots,X_n$ 是*相互独立的*（mutually independent），如果

$$f(x_1,x_2,\ldots,x_n)=f_1(x_1)f_2(x_2)\cdots f_n(x_n) \tag{18}$$

对于所有 $x_1, x_2, \cdots, x_n$ 都成立.

例 1.3.9

考虑例 1.3.8 中的试验，则 X（感觉更容易与母亲交流的女孩的数目）的概率函数为：

$$f_1(x) = P(X = x) = \binom{6}{x} p^x (1-p)^{6-x} \tag{19}$$

且 Y（感觉更容易与母亲交流的孩子总数）的概率函数为：

$$f_2(y) = P(Y = y) = \binom{14}{y} p^y (1-p)^{14-y} \tag{20}$$

由于

$$f(x, y) = P(X = x, Y = y) = P(X = x | Y = y) P(Y = y)$$

由（16）和（20）式，得 X 和 Y 的联合概率函数为：

$$f(x, y) = \frac{\binom{6}{x}\binom{8}{y-x}}{\binom{14}{y}} \binom{14}{y} p^y (1-p)^{14-y} = \binom{6}{x}\binom{8}{y-x} p^y (1-p)^{14-y}$$

但由于

$$f_1(x) f_2(y) = \binom{6}{x}\binom{14}{y} p^{x+y} (1-p)^{20-x-y}$$

所以，我们有

$$f(x, y) \neq f_1(x) f_2(y)$$

31 因此，X 和 Y 不独立. ■

习题

1. 如果 $f(x)$ 是二项分布概率函数，其中 $n = 6, p = 1/3$. 求：
   (a) $f(6)$.　(b) $f(0)$.　(c) $f(2.5)$.
   (d) $F(2.5)$.　(e) $F(-3)$.　(f) $F(7)$.
   (g) 画出概率函数的条形图.　(h) 画出分布函数图.
2. 假设 $f(x)$ 是离散均匀概率函数，其中 $N = 12$. 求：
   (a) $f(2)$.　(b) $f(12)$.　(c) $f(0)$.　(d) $f(1.5)$.
   (e) $F(0)$.　(f) $F(3.1)$.　(g) $F(1000)$.　(h) $F(-1000)$.
   (i) 画出概率函数的条形图.　(j) 画出分布函数图.
3. 假设 X 和 Y 独立，且服从二项分布的随机变量，X 对应的参数 $n = 3, p = 1/2$；Y 对应的参数 $n = 4, p = 1/2$. 设 $f(x, y)$ 是 X 和 Y 的联合概率函数，求：
   (a) $f(0,0)$.　(b) $f(0,1)$.　(c) $f(1,0)$.　(d) $f(3,4)$.
   (e) $f(4,4)$.　(f) $F(0,0)$.　(g) $f(1,1)$.　(h) $F(3,4)$.
4. 假设 $f(x)$ 是超几何分布概率函数，其中 $A = 3, B = 4$，求：

(a) 给定 $k=0$ 时，$f(0)$.　(b) 给定 $k=1$ 时，$f(1)$.　(c) 给定 $k=1$ 时，$f(2)$.
(d) 给定 $k=5$ 时，$f(1)$.　(e) 给定 $k=6$ 时，$f(1)$.

5. 一个用餐者从 6 种不同的三明治中随机地选取一种，
   (a) 样本空间是什么？
   (b) 样本空间上的概率函数是什么？
   (c) 在样本空间上定义一个服从离散均匀分布的随机变量.
6. 7 个男孩和 10 个女孩参加一场考试，假定每一个学生不及格的概率是 0.2，
   (a) 试验的样本空间是什么？
   (b) 已知有 3 个学生不及格，3 个学生都是男孩的概率是多少？
   (c) 你所采用概率分布的名字是什么？
   (d) 如果每个学生不及格的概率由 0.2 变为 0.8，(b) 的答案又是多少？ 32

思考题

1. 下面这些函数中的哪些可能是概率函数？并证明你的结论.
   (a) $f(x)=1/6$　当 $x=1,2,3,4$，
   $\quad\quad =0$　否则
   (b) $f(x)=(1/4)^x$　当 $x=1,2,3,4,\cdots$，
   $\quad\quad =0$　否则
   (c) $f(x)=(1-p)p^x$　当 $x=0,1,2,\cdots$，
   $\quad\quad =0$　否则，其中 p 是 $(0,1)$ 上的一个常数.
2. 假设每个患者都患有同种疾病，如果不经治疗的话，病人在一周之内痊愈的概率是 0.1. 现在向 10 个这种病人提供一种新药，一周之后 10 个病人中有 9 个痊愈了.
   (a) 若这种药物没有医疗作用，至少 9 人痊愈的概率是多少？
   (b) 从你的观点来看，你认为这种药有效吗？
   (c) 你所使用的样本空间是什么？
   (d) 在样本空间上你定义的概率函数是什么？
   (e) 在样本空间上你定义的随机变量是什么？
   (f) 你所采用随机变量概率分布的名字是什么？

1.4 随机变量的性质

我们已经讨论了随机变量的一些性质，比如概率函数和分布函数，分布函数描述了随机变量所有值得考虑的性质，因为分布函数揭示了随机变量所有可能的取值及其取值的概率. 然而，有时用整个分布函数来描述随机变量是不太方便或易混淆的，所以需要随机变量的某些“概括性描述”. 现在我们介绍随机变量的另外一些性质，它们可以用来对随机变量的分布进行简洁而不必完全的描述.

分位数

本书中常用来概括随机变量分布的方法就是给出随机变量某些特定的分位数

(quantile). “分位数”并不像“中位数”、“四分位数”、“十分位数”及“百分数”
33 那样为众人所知，后面的这些名词是比较常用的分位数. 例如，随机变量的中位数，是指随机变量取值比它大的概率不超过1/2，取值比它小的概率也不超过1/2的数. 把这个定义进行推广可得如下：

定义 1.4.1 对于(0,1)中的确定值 p，称 x_p 为随机变量 X 的 p 分位数 (the pth quantile of the random variable X)，如果 $P(X<x_p)\leqslant p$ 且 $P(X>x_p)\leqslant 1-p$.

如果 p 分位数的定义不唯一，为了避免混淆，采用惯例把所有满足定义1.4.1的数 x_p 中最大数与最小数的平均值取为其 p 分位数.

由定义可知，X 不超过 x_p 的概率不大于 p，X 大于 x_p 的概率不超过 $1-p$. 中位数(median)是0.5分位数，第三个十分位数(decile)是0.3分位数，上下四分位点(the upper and lower quartiles)分别是0.75和0.25分位数，第63个百分点(percentile)是0.63分位数.

也许寻找 p 分位数最简单直观的方法是使用随机变量的分布图. p 分位数就是在图上纵坐标是 p 的点所对应的横坐标，如下例.

例 1.4.1

设 X 是一个随机变量，有如下概率分布：

$$P(X=0)=1/4 \quad P(X=1)=1/4$$
$$P(X=2)=1/3 \quad P(X=3)=1/6$$

则 X 的分布函数可以用图1-4来表示. 0.75分位数 $x_{0.75}$，即上四分位点，可以用画
34 一条经过纵轴上点0.75的水平线来获得，如图1-4中的虚线所示，虚线与图相交点所对应的 x 的值就是上四分位数，本例中它为2. 故 $x_{0.75}=2$，也可以用定义来直接验证，因为

$$P(X<2)=1/2$$

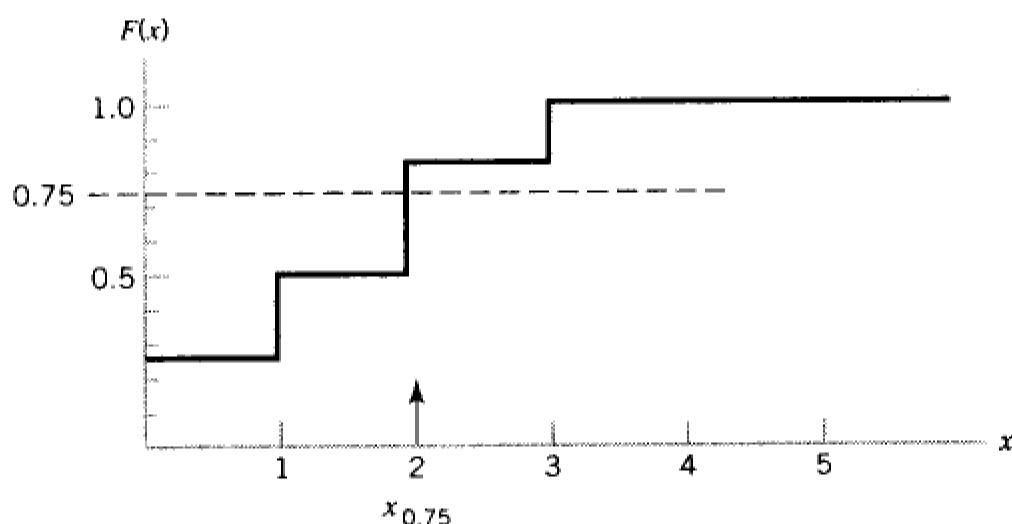


图1-4 z的分布函数

它小于0.75，且

$$P(X>2)=1/6$$

它小于 1 - 0.75.

中位数可类似地通过画一条经过纵轴上点 0.5 的线获得. 这里的中位数可取从 1 到 2 中的任意值，且容易看出这些值中的每一个都满足中位数的定义，通常我们取 1.5 作为中位数. ■

某些称之为“检验统计量”的随机变量在很多统计方法中起着非常重要的作用. 如果它们的分布函数完全未知，则这些统计量就没有什么用处. 附录中列出了很多表格，并给出了非参数统计方法中用到的一些检验统计量分布函数的信息. 分位数的使用浓缩了这些信息，否则这些表将非常臃肿.

通常定义随机变量后，研究的可能不是随机变量本身，而是这个随机变量的函数. 随机变量 X 的一个实值函数是代替 X 的取值而给样本空间中的点赋予新实值的准则，例如：$Y=X+4$，Y 是 X 的一个实值函数. 若 $X=x$，则 $Y=x+4$；若 $X=3$，则 $Y=7$，通常记作 $Y=u(X)$，此时，$u(X)$ 是 $X+4$. 其他的 $u(X)$ 也许是：$u(X)=X^2$，$u(X)=X$，或 $u(X)=(X-a)^2$，a 为某个常数. 尽管 Y 用 X 表示，由于 Y 也是对样本空间中的点赋予实值，所以我们认为 Y 也是一个随机变量. 一般地说，一个随机变量的实值函数也是一个随机变量.

期望值

随机变量的另一个很有用的量就是它的**期望值**（expected value），首先我们给出期望值的一般定义，然后再给出一些例子.

定义 1.4.2 设 X 是一个具有概率函数 $f(x)$ 的随机变量，且 $u(X)$ 是 X 的一个实值函数. 则 $u(X)$ 的期望值（记作 $E[u(X)]$）定义为：

$$E[u(X)]=\sum_x u(x)f(x) \tag{1}$$ 35

其中，求和是对 X 的所有可能值进行的. 如果（1）式右端的求和是无穷大，或不存在，我们就说 $u(X)$ 的期望不存在.

实际中，我们主要关心的是两个期望值，即 X 的均值和方差.

定义 1.4.3 设 X 是一个具有概率函数 $f(x)$ 的随机变量，X 的均值（通常记作 μ）是

$$\mu=E(X) \tag{2}$$

由（1）式，我们有：

$$\mu=E(X)=\sum_x xf(x) \tag{3}$$

(3) 式表明均值与物理学中的“重心”是一样的，都是一个中心点，平衡点. 如果在某标尺上的每一个 X 取值处放置相应的正比于概率的重量，则这个标尺将恰好在均值处达到两边平衡. 正因为均值的这种可“查找”分布中心的位置的性质，均值也称作“位置度量”. 前面讨论过的均值和中位数是两种最常用的位置度量.

例 1.4.2

考虑具有“成功”和“失败”两个结果的简单试验，“成功”的概率是 p，“失败”的概率是 q（等于 $1-p$）. 如果“成功”发生，则令 $X=1$；若“失败”发生，则令 $X=0$. 这样 X 是 $n=1$ 的二项分布，由（3）式，X 的期望值为：

$$E(X) = 1(p) + 0(1-p) = p \tag{4}$$

亦即，X 的均值是 p. 如果发生的结果为等概率的，即 $p=1/2$，则 X 的均值也是 1/2. ■

例 1.4.3

考虑一个总在同一餐馆吃午餐的商人，餐馆的午餐有 4 种定价：4.00 美元、4.50 美元、5.00 美元和 5.50 美元. 凭以往的经验，他知道自己每一天选择 4.00 美元午餐的概率是
36 0.25，选择 4.50 美元午餐的概率是 0.35，选择余下两种价格午餐的概率都是 0.20. 设 X 为每天的午餐价格，以美元为单位，则 X 的概率函数是

$$P(X=4)=0.25 \quad P(X=4.5)=0.35$$
$$P(X=5)=0.20 \quad P(X=5.5)=0.20$$

由（3）式，得到 X 的均值为：

$$E(X) = (4)(0.25) + (4.5)(0.35) + (5)(0.20) + (5.5)(0.20) = 4.675$$

经过一段较长的时间后，商人发现，尽管他的每一次午餐花费不尽相同，但他的午餐平均支出却大约是 4.675 美元. ■

刻度

正如称均值和中位数为位置度量一样，度量随机变量的伸展或变异的量称为“刻度度量”. 一个基于分位数的刻度度量是*四分位极差*（interquartile range），它是由 $x_{0.75}$ 减去 $x_{0.25}$ 得到的. 另一个直接基于概率函数的刻度度量是极差，它等于随机变量的最大可能值减去最小可能值. 最常用的刻度度量是*标准差*（standard deviation），它等于如下所定义的方差的平方根.

定义 1.4.4 设 X 是一个有均值 μ 和概率函数 $f(x)$ 的随机变量，X 的方差（记作 σ^2 或 $\mathrm{Var}(X)$）定义为：

$$\sigma^2 = E[(X-\mu)^2] \tag{5}$$

由（1）式，X 的方差可以写作：

$$\begin{aligned}\sigma^2 &= \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x) \\ &= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)\end{aligned} \tag{6}$$

因为 $\sum_x f(x)=1$，且由（3）式，则（6）式变为：

$$\sigma^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2 \tag{7}$$

37 这是计算方差更常用的形式.

方差的正平方根称为 X 的*标准差*（standard deviation），记作 σ.

例 1.4.4

若 X 服从 $n=1$ 的二项分布，且 $P(X=1)=p, P(X=0)=1-p$. 在例 1.4.2 中求得 X 的均值为 p，因此由（6）式得

$$\sigma^2 = (1-p)^2(p) + (0-p)^2(1-p) = p(1-p) = pq \tag{8}$$

另外，我们还可用（7）式来计算 σ^2. 先用（1）式计算 $E(X^2)$

$$E(X^2) = (1)^2(p) + (0)^2(1-p) = p$$

则 X 的方差为

$$\sigma^2 = E(X^2) - \mu^2 = p - p^2 = p(1-p)$$

与（8）式一致. X 的标准差是 $\sqrt{p(1-p)}$. ■

例 1.4.5

有编号从 1 至 6 的 6 个相同筹码，一只猴子从中拿了一个，交给它的训练者. 样本空间是猴子取到的筹码，设 X 表示筹码上的数，若每一个筹码等可能被选中，其概率是 1/6，那么 X 服从离散均匀分布，则 X 的均值如下：

$$E(X) = 1(\tfrac{1}{6}) + 2(\tfrac{1}{6}) + 3(\tfrac{1}{6}) + 4(\tfrac{1}{6}) + 5(\tfrac{1}{6}) + 6(\tfrac{1}{6}) = 3\tfrac{1}{2}$$

X^2 的期望值如下：

$$E(X^2) = 1(\tfrac{1}{6}) + 4(\tfrac{1}{6}) + 9(\tfrac{1}{6}) + 16(\tfrac{1}{6}) + 25(\tfrac{1}{6}) + 36(\tfrac{1}{6}) = 15\tfrac{1}{6}$$

由（7）式，X 的方差为：

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = 15\tfrac{1}{6} - (3\tfrac{1}{2})^2 = 2\tfrac{11}{12}$$

标准差是方差的平方根，此例中它为 1.71. ■ 38

定义 1.4.2 中定义了单个随机变量的函数的期望值，这里可以推广到多个随机变量的联合函数的情况，它可以用来考虑两个随机变量的*协方差*（covariance），也能求几个随机变量和的均值和方差.

定义 1.4.5 设随机变量 $X_1, X_2, \cdots, X_n$ 联合概率函数为 $f(x_1, x_2, \cdots, x_n)$，$u(X_1, X_2, \cdots, X_n)$ 是 $X_1, X_2, \cdots, X_n$ 的实值函数，则 $u(X_1, X_2, \cdots, X_n)$ 的*期望值*（expected value）定义为：

$$E[u(X_1, X_2, \ldots, X_n)] = \sum u(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n) \tag{9}$$

其中，求和是对所有可能取到的 $x_1, x_2, \cdots, x_n$ 进行的.

$X_1, X_2, \cdots, X_n$ 的一个简单函数是

$$Y = X_1 + X_2 + \cdots + X_n \tag{10}$$

即随机变量 Y 的每一个值都与包含 X_i 试验的联合试验有关，它是所有 X_i 的取值之和.

$$\begin{aligned} E(Y) &= \sum (x_1 + \cdots + x_n) f(x_1, \ldots, x_n) \\ &= \sum x_1 f(x_1, \ldots, x_n) + \cdots + \sum x_n f(x_1, \ldots, x_n) \end{aligned} \tag{11}$$

其中，求和是对 $x_1, \cdots, x_n$ 所有可能的联合取值求和. 由定义 1.4.5 和（11）式立即得到

$$E(Y) = E(X_1) + \cdots + E(X_n) \tag{12}$$

这些计算结果可以陈述为如下定理：

定理 1.4.1 设 $X_1,X_2,\cdots,X_n$ 是随机变量，且

$$Y = X_1 + X_2 + \cdots + X_n$$

则有 $E(Y)=E(X_1)+E(X_2)+\cdots+E(X_n)$.

定理 1.4.1 中的结论在任何情形下都成立，不论随机变量独立与否. 对看上去很难求的一些随机变量和的均值问题，使用定理 1.4.1 后，常常使问题变得简单.

39 下面两个例题的结论将在后面的章节中用到.

例 1.4.6

设 Y 是 n 次独立基本试验中“成功”的总次数，每一个基本试验中，“成功”或者“失败”的概率分别为 p 或 $q=1-p$，则 Y 服从参数为 n,p 的二项分布. 实际上，Y 也可看作 n 个独立的随机变量 $X_1,X_2,\cdots,X_n$ 之和. 其中，$X_i=1$，若第 i 个基本试验的结果是“成功”；$X_i=0$，若第 i 个基本试验的结果是“失败”（i 从 1 到 n). 则

$$Y = X_1 + X_2 + \cdots + X_n$$

且由定理 1.4.1

$$E(Y) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

在例 1.4.2 中，X_i 的均值是 p，因此

$$E(Y) = np \tag{13}$$

即是二项分布的均值. ■

注意在二项分布中的基本试验认为是相互独立的，因此 X_i 是独立的. 这个假设在求均值时不需要用到.

在例 1.4.7 中，需要用到以下引理，它给出了连续整数和的简洁表达式.

引理 1.4.1

$$\sum_{i=a}^{N} i = \frac{(N+a)(N-a+1)}{2} \text{ 和 } \sum_{i=1}^{N} i = \frac{(N+1)N}{2}$$

证明 要求的和可用两种方式表达，设 $S=\sum_{i=a}^{N} i$，则

$$S = a + (a+1) + (a+2) + \cdots + (N-1) + N$$
$$S = N + (N-1) + (N-2) + \cdots + (a+1) + a$$

将上述两个等式相加，得到

$$2S = (N+a) + (N+a) + (N+a) + \cdots + (N+a) + (N+a)$$
$$= (N+a)(N-a+1)$$

因此

40 $$S = \sum_{i=a}^{N} i = \frac{(N+a)(N-a+1)}{2}$$

当 $a=1$ 时，得到

$$\sum_{i=1}^{N} i = \frac{(N+1)N}{2}$$

证毕.

例 1.4.7

罐中有 N 个筹码，编号从 1 到 N，从中依次取出 n 个，其中 n 小于 N，记下对应的编号，置于一旁. 设 Y 是取出的 n 个筹码上的编号之和，假设抽取是随机的，即每个筹码等可能取到.

Y 的均值如果不用定理 1.4.1 是不易求的. 由于一旦记录了一个筹码的编号，则其他的筹码就不会出现这个编号，故此连续抽取不是独立的. 然而，我们仍把 Y 作为随机变量 $X_1, X_2, \cdots, X_n$ 之和，其中 X_i 表示第 i 次取出筹码的编号，它有概率函数

$$P(X_i = k) = \frac{1}{N}, \text{其中 } k = 1, 2, 3, \ldots, N$$

因此，借助引理 1.4.1，我们有

$$E(X_i) = \sum_{k=1}^{N} k\left(\frac{1}{N}\right) = \frac{(N+1)}{2} \tag{14}$$

（14）式给我们提供了一个离散均匀分布的均值. 因为 $Y = X_1 + X_2 + \cdots + X_n$，所以我们有

$$E(Y) = E(X_1) + E(X_2) + \cdots + E(X_n) = n\frac{N+1}{2} \tag{15}$$

■

协方差

两个随机变量一个特别有用的函数是

$$[X_1 - E(X_1)][X_2 - E(X_2)]$$

它的期望值称为 X_1 与 X_2 的*协方差*（covariance）. 特别地，比较定义 1.4.4 与下面的定义可以发现，X_1 的方差就是它自身的协方差. 41

定义 1.4.6 设随机变量 X_1 和 X_2 分别具有均值 μ_1 和 μ_2，概率函数 $f_1(x_1)$ 和 $f_2(x_2)$，其联合概率函数为 $f(x_1, x_2)$. X_1 和 X_2 的*协方差*定义为：

$$\text{Cov}(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)] \tag{16}$$

由关于期望值的定义 1.4.5，可得：

$$\text{Cov}(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)] = \sum(x_1 - \mu_1)(x_2 - \mu_2)f(x_1, x_2) \tag{17}$$

求和是对所有 x_1, x_2 的可能取值进行的，展开得

$$\begin{aligned}\text{Cov}(X_1, X_2) &= \sum(x_1x_2 - \mu_1x_2 - \mu_2x_1 + \mu_1\mu_2)f(x_1, x_2)\\ &= E(X_1X_2) - \mu_1\mu_2 - \mu_2\mu_1 + \mu_1\mu_2\\ &= E(X_1X_2) - \mu_1\mu_2\end{aligned} \tag{18}$$

在计算协方差时，（18）式通常比（17）式更好用.

例 1.4.8

保险公司发现每个人在一年内发生一次交通事故的概率是 0.1，但如果知道他上一年发生过交通事故，那么这个概率变成了 0.3.

设 X_1 取值 0 或 1，分别对应于某人在他保险期间的第一年内没有发生事故以及至少

发生过一次事故；X_2 类似于 X_1 的定义，对应于他保险期间的第二年的情况，则 X_1 的概率函数是

$$P(X_1=0)=0.9 \quad P(X_1=1)=0.1$$

X_2 也有相同的概率函数. 从例 1.4.2，我们得到

$$E(X_1)=0.1 \quad E(X_2)=0.1$$

X_1 和 X_2 联合概率函数在 $X_1=1$ 和 $X_2=1$ 时的值，可由下式表示：

$$f(1,1)=P(X_1=1,X_2=1)=P(X_2=1|X_1=1)P(X_1=1)$$

42 $$=(0.3)(0.1)=0.03$$

由定义 1.4.5 可直接得 $E(X_1X_2)$ 的表达式如下：

$$E(X_1X_2)=(1)(1)f(1,1)+\text{“0”项}=0.03$$

再用（18）式，可得 X_1 和 X_2 的协方差为：

$$\mathrm{Cov}(X_1,X_2)=E(X_1X_2)-E(X_1)E(X_2)=0.03-(0.1)(0.1)=0.02$$

■

相关系数

现在我们来定义*相关系数*（correlation coefficient），它是两个随机变量间线性相关性的一种度量. 这里我们不加证明地陈述两个结论：相关系数总是在 -1 和 $+1$ 之间；当两个随机变量相互独立时，它一定等于 0（在其他情形也可能是 0）.

定义 1.4.7　两个随机变量的*相关系数*（correlation coefficient）是它们的协方差除以它们标准差的乘积. 相关系数常用 ρ 表示，由下式给出

$$\rho=\frac{\mathrm{Cov}(X_1,X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} \tag{19}$$

下面给出在例 1.4.9 中要用到的一个引理，这个引理提供了前 N 个连续整数的平方和的简洁表达式.

引理 1.4.2

$$\sum_{i=1}^{N} i^2=\frac{N(N+1)(2N+1)}{6}$$

证明　设 $S=\sum_{i=1}^{N} i^2$，则

$$\begin{aligned}
S &= 1^2+2^2+3^2+4^2+\cdots+N^2\\
&= 1+2+3+4+\cdots+N\\
&\qquad +2+3+4+\cdots+N\\
&\qquad\quad +3+4+\cdots+N\\
&\qquad\qquad +4+\cdots+N\\
&\qquad\qquad\quad \cdots\quad\cdots
\end{aligned}$$

43 $$+N$$

其中在第 i 列数的和是 i^2，可我们并不对列求和而是对行求和. 从第 j 行顶端开始，对第 j 行的所有数求和，由引理 1.4.1 得

$$j+(j+1)+(j+2)+\cdots+N=\frac{(N+j)(N-j+1)}{2}=\tfrac{1}{2}(N^2+N+j-j^2)$$

再把这些行加起来，得

$$\begin{aligned}S&=\sum_{j=1}^{N}\tfrac{1}{2}(N^2+N+j-j^2)\\&=\tfrac{1}{2}\sum_{j=1}^{N}N^2+\tfrac{1}{2}\sum_{j=1}^{N}N+\tfrac{1}{2}\sum_{j=1}^{N}j-\tfrac{1}{2}\sum_{j=1}^{N}j^2\\&=\tfrac{1}{2}(N\cdot N^2)+\tfrac{1}{2}(N\cdot N)+\tfrac{1}{2}\cdot\frac{(N+1)N}{2}-\tfrac{1}{2}S\end{aligned}$$

这里第 2 个等式中的最后一个求和项可由 S 表示. 因而，移项得到：

$$\tfrac{3}{2}S=\tfrac{1}{4}(2N^3+3N^2+N)=\tfrac{1}{4}N(N+1)(2N+1)$$

所以

$$S=\sum_{i=1}^{N}i^2=\frac{N(N+1)(2N+1)}{6}$$

证毕.

例 1.4.9

如例 1.4.7，一个罐中有编号从 1 到 N 的 N 个塑料筹码，试验是从罐中取出 n 个，其中 $n\leqslant N$. 假设每个筹码都是等可能取出的且不放回. 设 $X_1,X_2,\cdots,X_n$ 为随机变量，其中 X_i 表示第 i 次取出的筹码的编号，$i=1,2,\cdots,n$. 在本例中，我们要求 X_i 和 X_j 的协方差. 由例 1.4.7，我们得到 X_i 的均值

$$E(X_i)=\frac{N+1}{2}$$ 44

且由（7）式和引理 1.4.2，我们得 X_i 的方差

$$\begin{aligned}\mathrm{Var}(X_i)&=E(X_i^2)-[E(X_i)]^2=\sum_{k=1}^{N}k^2\frac{1}{N}-\left(\frac{N+1}{2}\right)^2\\&=\frac{1}{N}\,\frac{N(N+1)(2N+1)}{6}-\left(\frac{N+1}{2}\right)^2=\frac{(N+1)(N-1)}{12}\end{aligned}\tag{20}$$

下面我们求随机变量 X_i 和 X_j 的协方差，其中 $i\neq j$. 它们的联合概率函数为

$$\begin{aligned}f(x_i,x_j)&=P(X_i=x_i,X_j=x_j)=P(X_i=x_i|X_j=x_j)\cdot P(X_j=x_j)\\&=\frac{1}{N-1}\cdot\frac{1}{N}\quad 其中\ x_i,x_j=1,2,\ldots,N;\ x_i\neq x_j\end{aligned}\tag{21}$$

由定义 1.4.6，X_i 和 X_j 的协方差可如下表示：

$$\begin{aligned}\mathrm{Cov}(X_i,X_j)&=E\{[X_i-E(X_i)][X_j-E(X_j)]\}\\&=\sum_{\substack{k=1\\k\neq s}}^{N}\sum_{s=1}^{N}\left(k-\frac{N+1}{2}\right)\left(s-\frac{N+1}{2}\right)\frac{1}{(N-1)N}\end{aligned}$$

其中，求和是对所有的 k,s 从 1 到 N，且 $k \neq s$ 进行的（X_k 与 X_s 不能同时取相同的值）. 如果同时加上和减去 $k=s$ 的项，则方差为

$$\mathrm{Cov}\,(X_i,X_j)=\frac{1}{(N-1)N}\sum_{k=1}^{N}\left(k-\frac{N+1}{2}\right)\sum_{s=1}^{N}\left(s-\frac{N+1}{2}\right)-\frac{1}{N-1}\sum_{k=1}^{N}\left(k-\frac{N+1}{2}\right)^2\frac{1}{N} \quad (22)$$

为简化（22）式，我们注意到

$$\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)=0 \quad (23)$$

由方差的定义及（20）式，我们得

$$\mathrm{Var}\,(X_i)=\sum_{k=1}^{N}\left(k-\frac{N+1}{2}\right)^2\frac{1}{N}=\frac{(N+1)(N-1)}{12} \quad (24)$$

把（23）式和（24）式代入（22）式，得到：

$$\mathrm{Cov}\,(X_i,X_j)=-\frac{N+1}{12} \quad (25)$$

■

45 协方差的重要性质就是讨论在两个独立随机变量的情形下协方差的取值. 设 X_1 和 X_2 是两个独立的随机变量，分别有概率函数 $f_1(x_1)$ 和 $f_2(x_2)$，均值 μ_1 和 μ_2，则 X_1 和 X_2 的协方差为

$$\begin{aligned}\mathrm{Cov}\,(X_1,X_2)&=\sum_{x_1,x_2}x_1x_2f_1(x_1)f_2(x_2)-\mu_1\mu_2=\left[\sum_{x_1}x_1f_1(x_1)\right]\left[\sum_{x_2}xf_2(x_2)\right]-\mu_1\mu_2\\&=\mu_1\mu_2-\mu_1\mu_2=0\end{aligned}$$

这说明，两个随机变量的独立性意味着它们的协方差是 0，即它们的相关系数也是 0.

定理 1.4.2 如果 X_1 和 X_2 是两个独立的随机变量，则 X_1 与 X_2 的协方差是 0.

下面例 1.4.10 表明，定理 1.4.2 的逆命题不一定正确. 即，协方差是 0 并不意味着随机变量是独立的，而在实际中却经常出现这样的误解.

例 1.4.10

定义两个随机变量的联合概率函数如下

$$P(X=0,Y=0)=1/2$$
$$P(X=1,Y=1)=1/4$$
$$P(X=-1,Y=1)=1/4$$

则 X 的概率函数为

$$P(X=0)=1/2$$
$$P(X=1)=1/4$$
$$P(X=-1)=1/4$$

Y 的概率函数为

$$P(Y=0)=1/2 \quad P(Y=1)=1/2$$

X 和 Y 的期望值为

46 $$E(X)=0 \quad E(Y)=1/2$$

X 和 Y 的协方差为

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=(1)(\tfrac{1}{4})+(-1)(\tfrac{1}{4})-(0)(\tfrac{1}{2})=0$$

然而，X 和 Y 却不独立，因为

$$P(X=0,Y=0)=1/2$$

它不等于

$$P(X=0)P(Y=0)=(\tfrac{1}{2})(\tfrac{1}{2})=1/4$$

因此 X 和 Y 有 0 协方差，但它们却不独立. ■

下面我们准备求一些随机变量和的方差. 设 $Y=X_1+X_2+\cdots+X_n$，其中，X_i 之间可能独立也可能不独立. 要求 Y 的方差，由方差的定义有：

$$\mathrm{Var}(Y)=E\{[Y-E(Y)]^2\}=E\{[X_1+X_2+\cdots+X_n-E(X_1)-E(X_2)-\cdots-E(X_n)]^2\}$$
$$=E\left\{\sum_{i=1}^{N}[X_i-E(X_i)]^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq j}}^{N}[X_i-E(X_i)][X_j-E(X_j)]\right\}$$

但是，由于随机变量和的期望值等于这些随机变量期望值的和，所以

$$\mathrm{Var}(Y)=\sum_{i=1}^{n}E\{[X_i-E(X_i)]^2\}+\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}E\{[X_i-E(X_i)][X_j-E(X_j)]\}$$
$$=\sum_{i=1}^{n}\mathrm{Var}(X_i)+\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\mathrm{Cov}(X_i,X_j) \tag{26}$$

如果 $X_1,X_2,\cdots,X_n$ 相互独立，由定理 1.4.2，我们有 $\mathrm{Cov}(X_i,X_j)=0$，且

$$\mathrm{Var}(Y)=\sum_{i=1}^{N}\mathrm{Var}(X_i) \tag{27}$$

我们把上述内容总结成如下定理： 47

定理 1.4.3 设 $X_1,X_2,\cdots,X_n$ 是随机变量且

$$Y=X_1+X_2+\cdots+X_n$$

则

$$\mathrm{Var}(Y)=\sum_{i=1}^{n}\mathrm{Var}(X_i)+\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\mathrm{Cov}(X_i,X_j)$$

进一步讲，如果 $X_1,X_2,\cdots,X_n$ 相互独立，则

$$\mathrm{Var}(Y)=\sum_{i=1}^{n}\mathrm{Var}(X_i)$$

例 1.4.11

续例 1.4.9. 设 X_i 是取出的第 i 个筹码的编号，且 Y 是所有 X_i 的和（见例 1.4.7），则由定理 1.4.3 得到

$$\mathrm{Var}(Y)=\sum_{i=1}^{n}\mathrm{Var}(X_i)+\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\mathrm{Cov}(X_i,Y_j)$$

等式中的诸项由（20）式和（25）式给出，其中方差项出现了 n 次，协方差项出现

了 $n(n-1)$ 次，亦即

$$\mathrm{Var}(Y)=n\frac{(N+1)(N-1)}{12}+n(n-1)\left(-\frac{N+1}{12}\right)=\frac{n(N+1)(N-n)}{12} \tag{28}$$

注意，$\mathrm{Var}(X_i)$ 是 $n=1$ 时 $\mathrm{Var}(Y)$ 的特殊情形. ■

二项分布的协方差在下面的例 1.4.12 中给出.

例 1.4.12

考虑 n 次独立的基本试验，每个基本试验结果是“成功”或“失败”，且“成功”和“失败”的概率分别是 p 和 q，其中 $p+q=1$. 与例 1.4.4 和例 1.4.6 类似，设 X_i 取 0 或 1，分别对应于第 i 个基本试验是“失败”或“成功”，设 Y 为 n 次试验中总的“成功”次数. 由 X_i 相互独立及定理 1.4.3，有

48

$$\mathrm{Var}(Y)=\sum_{i=1}^{n}\mathrm{Var}(X_i)$$

由例 1.4.4 得，$\mathrm{Var}(X_i)=pq$，所以

$$\mathrm{Var}(Y)=npq$$

这就得到了二项分布 Y 的方差. ■

前面例题的一些结果在这本书的后面要用到，为方便起见，把它们陈述为相应的定理.

定理 1.4.4 设 X 是服从二项分布的随机变量

$$P(X=k)=\binom{n}{k}p^kq^{n-k}$$

则 X 的均值和方差由下式给出

$$E(X)=np \quad \mathrm{Var}(X)=npq$$

定理 1.4.5 设 X 是从 1 至 N 这 N 个整数中，无放回地随机取出的 n 个整数的和，则 X 的均值和方差是

$$E(X)=\frac{n(N+1)}{2} \quad \mathrm{Var}(X)=\frac{n(N+1)(N-n)}{12}$$

例 1.4.13

广告代理商为他们的一位客户挑选了 12 个样品杂志广告，并把这些广告按照他们认为的对出售商品的影响力大小进行排序. 最有效的广告排序为 1，等等. 这名客户（产品的生产者）选择购买了 4 种广告，代理商对它们的排序是 4,6,7 和 11.

假设客户的选择和代理商的排序是独立的，则被选广告排序和的分布等同于与从标号 1 到 12 的 12 个筹码中取出 4 个筹码标号和的分布. 设 X 为被选的 4 个广告的排序和，由于被选的广告独立于排序，由定理 1.4.5 可得 X 的均值为

$$E(X)=\frac{(4)(12+1)}{2}=26$$

且 X 的方差为

$$\mathrm{Var}(X)=\frac{(4)(12+1)(12-4)}{12}=34\tfrac{2}{3}$$ 49

X 的标准差为

$$\sigma=\sqrt{\mathrm{Var}(X)}=5.9$$

X 的观测值为

$$X=4+6+7+11=28$$

在前面的假设下，这个值与 X 的均值比较接近. ■

习题

1. 若 $P(X=0)=1/3, P(X=1)=1/3, P(X=2)=1/6$，且 $P(X=3)=1/6$，求
   (a) $E(X)$. (b) $\mathrm{Var}(X)$. (c) $E(X^2+2X)$.
   (d) 中位数. (e) $x_{1/3}$. (f) 第 4 个十分位数.
2. 若 $P(X=0)=0, P(X=1)=1/2, P(X=2)=1/4, P(X=4)=1/4$，求
   (a) $E(X)$. (b) $\mathrm{Var}(X)$. (c) $E(-X)$.
   (d) 中位数. (e) 上四分位数. (f) 第 37 百分位数.
3. 若 $P(X=0,Y=0)=1/4, P(X=0,Y=1)=1/4, P(X=1,Y=0)=1/4$，且 $P(X=1,Y=1)=1/4$，求
   (a) $E(X)$. (b) $E(Y)$. (c) $E(XY)$.
   (d) $E(X+Y)$. (e) $\mathrm{Cov}(X,Y)$. (f) $P(X=0)$.
   (g) $P(X=1)$. (h) X 与 Y 独立吗?
4. 若 $P(X=0,Y=0)=1/8, P(X=0,Y=1)=3/8, P(X=1,Y=0)=3/8$，且 $P(X=1,Y=1)=1/8$，求
   (a) $E(X)$. (b) $E(Y)$. (c) $E(XY)$.
   (d) $E(X^2Y)$. (e) $\mathrm{Cov}(X,Y)$. (f) $P(X=x)$ 对每一 x.
   (g) X 与 Y 独立吗?
5. 从 1 到 66，这 66 个整数的和是多少?
6. 从 70 到 99，这 30 个整数的和是多少?
7. 若 X 为投掷一次均匀骰子得到的点数，求
   (a) $E(X)$. (b) $\mathrm{Var}(X)$. (c) $E(X^2+X)$. 50
8. 若从 1 到 30 连续地对 30 张票编号，从中无放回地随机取出 2 张票，求
   (a) E（这两张票的编号和). (b) Var（这两张票的编号和).
9. 若 10 个顾客从 1 到 10 连续地编号，随机地选出 2 个进行访问. 求这两位顾客编号和的均值、方差和极差.
10. 若 100 个顾客从 1 到 100 连续地编号，随机地选出 12 个进行访问. 求选中顾客编号和的均值、方差和极差.
11. 若 $P(X=0,Y=0)=1/3, P(X=0,Y=1)=1/3$，且 $P(X=1,Y=0)=1/3$.
    (a) 求 X 的边缘概率分布. (b) 求 $E(X^2Y)$.
    (c) 求 $\mathrm{Cov}(X,Y)$. (d) 求 X,Y 的相关系数.
12. 二维随机变量 (X,Y) 取值 (1,1) 的概率是 0.25，取值 (2,1) 的概率是 0.25，取值

(1,2) 的概率是0.50. 求 X 的中位数，Y 的方差和 X、Y 的协方差. 假设 $F(x,y)$ 是 (X, Y) 的分布函数，求 $F(1,2)$ 并画出 X 的分布函数及 Y 的概率函数图.

13. 有8匹马参加赛马大赛，其中有3匹马来自 Lubbock，它们的位置（1至8）是无放回地随机指定的. 设 X 表示来自 Lubbock 的那3匹马的位置编号和，求 X 的均值和方差.
14. 设 X 与 Y 是两个独立的二项分布随机变量，对于 $X, n=3, p=1/3$；对于 $Y, n=4, p=1/4$. 求 $F(1,1)$ 以及 $E(XY)$.

思考题

1. 在随机变量 X 的均值处画一条竖直线，它把 X 的分布函数进行了划分. 证明：在分布函数下方、0上方并在均值左边的区域的面积等于在分布函数上方、1下方并在均值右边的区域的面积（提示：画一个分布函数图，考虑要研究的面积）.
2. 用由引理1.4.1获得引理1.4.2的类似方式，用引理1.4.2来证明如下的引理1.4.3

$$\sum_{i=1}^{N} i^3 = \frac{N^2(N+1)^2}{4}$$

51 [更一般的推广式由 Iman（1970）给出.]

1.5 连续型随机变量

到目前为止，本章中所介绍的所有随机变量有一个共同的特征：它们可能的取值是可列的. 二项分布随机变量的可能取值是 $0,1,2,3,4,\cdots,n-1,n$；而其他值则取不到. 离散均匀分布的可能取值是 $1,2,3,\cdots,N$. 对于在前面定义和例题中引入的随机变量，类似的可列值都可以写出.

这些可列值可能是无限长的，比如在一个试验中，随机变量 X 等于一只猴子在最后按"对"钮且拿到奖品之前按"错"钮的次数，那么，如果第一次按的就是"对"钮，则 $X=0$，或者如果这只猴子找到"对"钮比较困难，X 也可能等于1000. 理论上，这只猴子在最后按"对"钮之前选择按"错"钮的次数是没有上限的. 尽管这种可列值会无限地长，我们还是可以列出 X 的所有可能取值. 无限长的可列值是这个模型的一个特征，而并非这个例子和很多情况下的实际试验真得如此，因为很多实际因素，比如说猴子的最终死亡、研究经费的缺乏或者试验者的试验热情的减少，都会使实际试验不可能延长到近似荒谬的那一步. 然而，模型是合理的，且模型中的随机变量可以有无穷多的取值.

离散型随机变量

一个随机变量的可能取值是可列的，更确切表述方式是：随机变量的可能取值与部分或全部正整数之间存在着一一对应（one-to-one correspondence）. 这意味着对随机变量的每一个可能取值存在唯一的正整数与它对应，且这个正整数不与随机变量除此之外的其他可能值相对应. 具有这种性质的随机变量称为是离散型

随机变量. 目前我们考虑的所有随机变量都是离散的，尽管如此，前面已证明过的一些定理却适用于所有的随机变量，即使我们仅对离散型随机变量证明了它们.

定义 1.5.1 称一个随机变量 X 是离散（discrete）的随机变量，如果 X 的可能取值与部分或全部自然数之间存在着一一对应关系.

离散型随机变量的分布函数总是一个阶梯函数（step function），即它的图形看上去像楼梯的一串台阶，尽管这些台阶可能会不均匀，甚至可能会有无穷多步台阶. 52
如果这个图形的某一部分是逐渐连续上升，而不是呈阶梯上升的，那么相应的随机变量就不是离散的.

连续型随机变量

如果一个分布函数没有阶梯，上升的地方都是逐渐连续上升的，则称这个分布函数是连续的（continuous），与这个分布函数相对应的随机变量称为连续型随机变量（continuous random variable）. 图 1-5 是一个连续型分布函数的图形.

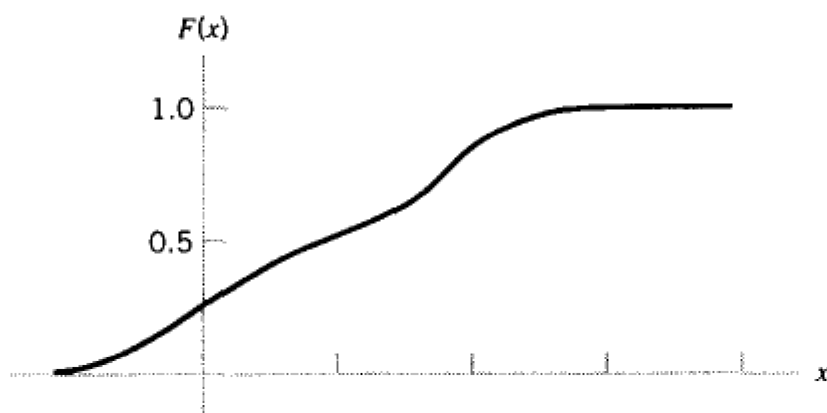


图 1-5 连续型分布函数的图形

说一个分布函数没有阶梯，即是说不存在这样的两条水平线，它们与图形的交点在水平轴上的值是相同的. 也就是说，若一个分布函数存在阶梯，则至少可以作出这样的两条线，比如说高是 p_1 和 p_2，两者非常接近以至于它们与分布函数图形的交点在水平轴上的值是相同的. 这实际上是描述了找分位数的图形法，可以说在分布函数中若有一个阶梯，那么至少存在两个彼此相同的分位数 x_{p_1} 和 x_{p_2}；相反地，如果没有精确相等的两个分位数，那么这个分布函数就没有阶梯，即函数是连续的. 这就引出连续型随机变量的如下定义.

定义 1.5.2 称一个随机变量 X 是连续型随机变量，如果不存在 X 的两个相等的分位数 x_{p_1} 和 x_{p_2}，其中 $p_1 \neq p_2$. 或等价地，称 X 是连续型随机变量，如果对所有的数 x, $P(X \leq x)$ 等于 $P(X < x)$.

例 1.5.1

图 1-6 中的分布函数是连续型的分布函数，且任何一个以 $F(x)$ 为分布函数的随机变量都是连续型随机变量. 典型的连续型随机变量包括时间、高度、距离、体积的测量等等. ■ 53

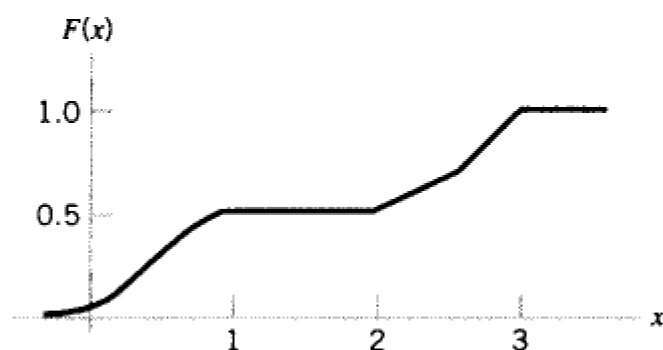


图1-6 连续型分布函数

事实上，实际的随机变量没有连续的，因为它们观测值都是某些测量的结果，测量工具在区分两个值时的能力也是有限的. 连续型随机变量可以在理论上存在，如在一个实际试验的模型中. 而有时即使知道随机变量是离散的，我们也倾向于假定它是一个连续型随机变量，如下面的例1.5.2.

例1.5.2

赛马跑完一英里赛程所需的时间是一个连续的量，因为时间一般是一个连续的量. 然而在实际中，时间是以不到1/5秒为最小测量单位的，一匹马跑完两场比赛所用的时间（即测量时间）相同是不太常见的. 时间长度确切相等的概率是0. 因此把比赛的精确时间假设成连续的是合理的，这个时间近似等于比赛的测量时间，而比赛的测量时间是离散型随机变量. 如果两匹马在同一场比赛中，领先于其他所有的马以相同的测量时间跑过终点线，那么比赛的胜者将通过两匹马冲过终点线时所拍摄的照片来分析决定. 这样来确定实际的获胜者很少会失误，尽管赛马比赛可能会出现两个或多个测量时间上看上去相同，但比赛的次序仍旧可以确定，这说明即使随机变量（测量时间）是离散的，也可以作为实际连续时间的一种近似. ■

考虑连续型随机变量的另一原因是，离散型随机变量的分布函数有时可以用连续型的分布函数来逼近，这样，离散型随机变量中的有关简单方法就可用来近似计算所要的一些概率结果. 常用的两个连续型分布是正态分布（normal distribution）和
54 χ^2分布（chi-squared distribution，读作卡方分布）.

正态分布

下面分布函数的定义或许会使那些对初等微积分不太了解的人感到吃惊. 然而，这无需担忧，因为我们已制定好了它的分布函数表，这样的表（如表A1）在大多数统计教材中都可以找到；不仅如此，大多数电脑里的统计程序或软件需要时都可以计算分位数和概率.

定义1.5.3 设X是一个随机变量，称X服从正态分布，如果它的分布函数具有如下形式：

$$F(x)=P(X\leqslant x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-1/2[(y-\mu)/\sigma]^2}dy \quad (1)$$

其中，可以证明（用微积分）参数μ和σ分别是X的均值和标准差. 标准正态分布

(standard normal distribution) 是对应 $\mu=0,\sigma=1$ 的正态分布.

正态分布函数的值不能直接计算出，所以表 A1 可用来近似计算与正态分布有关的概率. 表 A1 给出了标准正态分布的1000 个分位数. 利用下面这个未证明的定理中的等式，我们可在表 A1 中找到均值为μ、方差为σ^2 的正态分布的分位数.

定理 1.5.1 对于给定的值 p，设 x_p 是均值为μ、方差为 σ^2 的正态分布的 p 分位数，z_p 是标准正态分布的 p 分位数，则分位数 x_p 可通过 z_p 由如下线性关系获得：

$$x_p=\mu+\sigma z_p \tag{2}$$

类似地，z_p 也可通过 x_p 由如下线性关系获得：

$$z_p=\frac{x_p-\mu}{\sigma} \tag{3}$$

例 1.5.3

设 Z 服从标准正态分布，使用表 A1 来计算 Z 不超过 1.42 的概率. 由表 A1，我们得到 55

$$P(Z\leqslant 1.4187)=0.922$$

且

$$P(Z\leqslant 1.4255)=0.923$$

我们简单地用接近 1.42 的分位数得到

$$P(Z\leqslant 1.42)\cong 0.922$$

■

例 1.5.4

从一群人中随机选取一个人的 IQ，设为 X，假设 X 是均值为 100，标准差为 15 的正态分布. 我们要计算 X 大于 125 的概率. 由于

$$P(X>125)=1-P(X\leqslant 125)$$

这就归结于求 P ($X\leqslant125$)，对应于分位数 $x_p=125$，由 (3) 式可以得到标准正态分布的分位数 z_p，

$$z_p=\frac{x_p-\mu}{\sigma}=\frac{125-100}{15}=1.6667$$

再由表 A1 可以看到，若 $z_p=1.6667$，则 $p=0.952$. 因此 125 是 X 的 0.952 分位数.

$$P(X\leqslant 125)=0.952 \quad P(X>125)=0.048$$

故所求的概率是 0.048.

为了获得上 1 百分位数，也称作第 99 个百分位数，我们需要求 $x_{0.99}$，其中

$$P(X\leqslant x_{0.99})=0.99$$

这样由表 A1，$z_{0.99}=2.3263$，$x_{0.99}$可由 (2) 式得到

$$x_{0.99}=\mu+\sigma z_{0.99}=100+15(2.3263)=135$$

因此，随机选取这人的 IQ 小于 135 的概率是 0.99. ■

例 1.5.5

铁路公司观察了一段时间发现，乘某一列车的人数X似乎服从均值为540，标准差 56

为32的正态分布. 如果想以95%的概率保证每位乘客都有座位，那么公司应该在这辆列车上提供多少个座位呢?

实际上，我们要计算第95个百分位数. 由 (2) 式可得

$$x_{0.95}=\mu+\sigma z_{0.95}=540+32(1.6449)=593$$

其中，$z_{0.95}$由表A1得到，即这个公司在这辆列车上需要提供593个座位，才能以95%的概率保证，在列车的任何一次运行中每一个乘客都有座位. ■

例1.5.5中随机变量X实际上是一个取非负整数的离散型随机变量，严格地说，X不可能服从正态分布，而且对X的分布也很难找到一个理想的离散分布，因此对它做正态逼近一部分是为了简便，也有一部分是出于需要. 在其他的问题中，或许知道一个与数据吻合得较好的离散分布，但为了计算方便，我们还是会使用正态逼近. 使用正态逼近的理论根据通常由中心极限定理来保证.

中心极限定理

所谓的中心极限定理可能有多种形式，所有的形式有一个共同的特点，就是在一定的条件下，一些随机变量的和能够用正态分布逼近. 这个定理说：随着被求和的随机变量的个数的增大（即趋于正无穷），并在满足其他一般条件的情况下，一些随机变量的和将趋于正态分布. 这些“其他的一般条件”可以有多种描述方式，由此产生了不同形式的中心极限定理. 对这个定理的详细讨论已超出了本书的范围.

定理1.5.2（中心极限定理，Central Limit Theorem） 设Y_n是n个随机变量$X_1,X_2,\cdots,X_n$的和，μ_n是Y_n的均值，σ_n^2是Y_n的方差. 在一般较容易满足的条件下，当随机变量的个数n趋于无穷大时，下面随机变量的分布函数

$$\frac{Y_n-\mu_n}{\sigma_n}$$

趋于标准正态分布函数.

57 在实际中，求和的随机变量的个数不会达到无穷大，但是中心极限定理的价值在于，在定理成立的情形中，只要n比较“大”，正态逼近总是“相当地好”，这里“相当地好”和比较“大”都是主观的说法. 所以在实际应用正态逼近时，有很多说法和观点. 通常如果n大于30时，正态逼近就比较满意了，但有时当n像5或10一样小时，正态逼近还是相当好的.

下面例1.5.6将说明正态逼近在二项分布中很好的应用.

例1.5.6

设Y_n是一个服从均值为np、方差为npq（见定理1.4.4）二项分布的随机变量（见定义1.3.5），则Y_n可以看作n个相互独立但都服从$n=1$的二项分布的随机变量之和（参见例1.4.12）. 对较大的n，随机变量

$$\frac{Y_n-np}{\sqrt{npq}}$$

的分布近似服从标准正态分布. 等价地说，对较大的 n, Y_n 的分布函数能够用均值为 np、方差为 npq 的正态分布来逼近（定理 1.5.1）. ■

下面的例子说明中心极限定理的另一个应用，在第 5 章中还会用到它.

例 1.5.7

考虑一个抽样方案，从 1 至 N 这 N 个整数中无放回地随机抽取 n 个整数. 设 X_i 是第 i 次取出的整数，且

$$Y_n = X_1 + X_2 + \cdots + X_n$$

是 n 个选出整数的和. 对于较大的 n 和 N，下式随机变量的分布函数

$$\frac{Y_n - \dfrac{n(N+1)}{2}}{\left(\dfrac{n(N+1)(N-n)}{12}\right)^{\frac{1}{2}}}$$

可以用标准正态分布函数来逼近（定理 1.4.5）. 换句话说，Y_n 的分布函数能够用均值为 $n(N+1)/2$、方差为 $n(N+1)(N-n)/12$ 的正态分布函数来逼近（定理 1.5.1）. ■ 58

χ^2 分布

中心极限定理之所以成为非常有用的定理，是因为它具有广泛应用，同时在某种程度上，它也说明了正态逼近的用处，所以正态分布是一个十分有用的分布. 其他和正态分布有关的分布由此也变得很重要，比如说 χ^2 分布.

下面定义的 χ^2 分布函数中用到了微积分中的“积分”符号和“Gamma 函数” $\Gamma(k/2)$. 这些符号不需要解释甚至不用理解，因为 χ^2 分布函数的值已制成了表（见表 A2），无论什么时候需要用到分布函数的值，你都可以使用它. 电脑里的很多统计程序和软件需要时都可以计算 χ^2 分布分位数和概率.

定义 1.5.4 随机变量 X 称为服从自由度为 k 的 χ^2 分布（chi-squared distribution with k degrees of freedom），如果 X 的分布函数满足

$$\begin{aligned} F(x) = P(X \leqslant x) &= \int_0^x \frac{y^{(k/2)-1} e^{-y/2}}{2^{k/2}\Gamma(k/2)} dy \quad \text{若 } x > 0 \\ &= 0 \quad \text{若 } x \leqslant 0 \end{aligned} \tag{4}$$

(4) 式的分布函数表明，χ^2 分布的随机变量只能取非负值，因为若 x 为负，则 $F(x)=0$. 自由度 k 是一个参数，它的选取一般局限于 1,2,3 等整数. 对于不同的参数 k，分布函数也是不同的. 表 A2 给出了一些 χ^2 随机变量选定的分位数，k 取 1,2,3，直到 30，也有部分大于 30 的 k 值. 当 k 大于 100 时，可以通过中心极限定理来获得近似的分位数，本节中的后面部分将给予介绍.

在一些数理统计的入门书中都列有结果：如果 X 是一个服从自由度为 k 的 χ^2 分布的随机变量，那么 X 的均值和方差分别为：

$$E(X)=k \quad (5)$$
$$\mathrm{Var}(X)=2k \quad (6)$$

下面这个定理的证明可在 Freund(1962, p. 194)中找到.

定理 1.5.3 设 $X_1, X_2, \cdots, X_n$，是 k 个独立同分布的标准正态随机变量，Y 是 X_i 的平方和

$$Y = X_1^2 + X_2^2 + \cdots + X_k^2 \quad (7)$$

59 那么 Y 服从自由度为 k 的 χ^2 分布.

例 1.5.8

一位儿童心理学家询问了 100 个儿童，了解他们更喜欢玩两辆玩具卡车中的哪一辆，这两辆卡车一个是红色的，另一个是绿色的，其他方面一模一样. 这位心理学家关心的是儿童对颜色是否有偏好.

令随机变量 X 等于选择绿色卡车的儿童数，结果 42 个儿童选择了绿色的卡车，其余 58 个选择了红色的卡车. 这个模型中假设是"没有偏好"，所以 X 应服从均值为 $np=50$，方差 $npq=25$ 的二项分布，这时我们可以考虑用 X 分布函数的正态逼近，所以

$$\frac{X-50}{5}$$

近似于标准正态随机变量. 然而，心理学家关心的是两个方面的决定性差异，即她想知道 X 是否比 50 小很多或大很多，所以本质上她用了偏差的平方，即实际考察的是随机变量

$$X^* = \left(\frac{X-50}{5}\right)^2$$

因为它可以与自由度为 1 的 χ^2 随机变量进行比较. 在这个试验中，$X^* = [(42-50)/5]^2 = 2.56$. 在表 A2 中使用内插法，且 $k=1$ 时，可得到一个比 2.56 小的值(相应于一个接近 50 的 X 的值) 的概率约为 0.88. 所以心理学家得出结论：这里的儿童有对颜色偏好的倾向. (这种得出结论的方法在后面的章节中将进行详细讨论.) ■

在例 1.5.8 中，随机变量$(X-50)/5$ 的分布函数近似于标准正态分布函数，所以 X^* 的分布近似于自由度为 1 的 χ^2 分布，所求的概率可由计算正态分布函数的上下尾概率获得，即

$$P(X^* \leqslant (-1.6)^2) = P\left(-1.6 < \frac{X-50}{5} < +1.6\right)$$

由表 A2 可得到下尾的概率，应该等于由表 A1 得到的上尾概率，两个概率间的唯一不

60 同在于两表中使用的内插法. 如果自由度大于 1，那么表 A1 就不能用来替代表 A2.

例 1.5.9

续例 1.5.8. 心理学家有两部相同的玩具电话，一部是白色的，另一部是蓝色的. 她让 25 个儿童选择一部玩，其中有 17 名儿童选择了白色的电话，另外 8 名选择了蓝色的. 设 Y 等于选择白色电话儿童数的随机变量，因为在没有颜色偏好的假设下，

$$\frac{Y-np}{\sqrt{npq}}=\frac{Y-(1/2)(25)}{5/2}$$

近似于一个标准正态随机变量. 随机变量

$$Y^*=\left(\frac{Y-(1/2)(25)}{5/2}\right)^2$$

可与自由度为 1 的χ^2 随机变量进行比较. 由于 $Y=17$, 那么 $Y^*=3.24$, 在表 A2 中运用内插法, 可求得自由度为 1 的χ^2 分布随机变量小于 3.24 的概率约为 0.92. 因此, 如果每一个玩具等可能选中的假设是合理的, 那么出现偏离期望值 12.5 这么大偏差的概率大约只有 8%.

因为例 1.5.8 和本例中的试验都是为同一个目的设计的, 那么, 应该通过某一方式来组合这些结果. 一个合理的想法是将 X^* 和 Y^* 看作相互独立的随机变量, 对于 X^* 和 Y^* 的下列组合

$$W=X^*+Y^*$$

利用定理 1.5.3, 则 W 的分布函数可以由自由度为 2 的χ^2 分布函数来逼近. 所以

$$W=2.56+3.24=5.80$$

由表 A2 和内插法可得, 自由度为 2 的χ^2 随机变量大于 5.80 的概率仅为 0.06.

在这个例子中, 通过结合两项研究中的信息来获得更多关于儿童中存在颜色偏好的信息.

需要注意的是: 若定义 Y 为偏好蓝色电话的儿童数 (代替本例中的 Y), 则 Y^* 的取值不变, 因为 Y 对均值的偏差取了平方, 所以消除了差异的方向影响. ■

例 1.5.9 中, 两个近似χ^2 的随机变量相加, 且它们的和近似于一个自由度为 2 的χ^2 随机变量. 一般来说, 这种组合独立χ^2 随机变量的方法是可行的, 有关这种方法更多的讨论可参见 Radhadkrishna(1965) 和 Nelson(1966). 61

下面的定理可在 Freund(1962, p194) 中找到.

定理 1.5.4 假设 $X_1,X_2,\cdots,X_n$ 分别是自由度为 $k_1,k_2,\cdots,k_n$ 的独立χ^2 随机变量. W 记为 X_i 的和, 则 W 为自由度为 k 的χ^2 随机变量, 其中

$$k=k_1+k_2+\cdots+k_n$$

本书后面将应用定理 1.5.4 来逼近几个随机变量和的分布函数, 其中, 随机变量假设相互独立且都近似于χ^2 随机变量.

由于自由度为 k 的χ^2 随机变量可认为是 k 个独立且服从自由度为 1 的χ^2 分布随机变量的和, 所以它们满足中心极限定理中的条件. 由 (5) 和 (6) 式可得到自由度为 k 的χ^2 随机变量的均值和方差分别为 k 和 $2k$. 因此, 如果 W 是自由度为 k 的χ^2 随机变量, 则当 k 较大时, 下式

$$Z=\frac{W-k}{\sqrt{2k}} \tag{8}$$

的分布函数逼近于标准正态分布函数. 由定理 1.5.1 可知, 若 z_p 是来自表 A1 的分位

数，对于较大的 k，相应于表 A2 中的分位数 w_p，可由下式来近似

$$w_p = k + \sqrt{2k}\, z_p \tag{9}$$

注意到，近似式（9）没有表 A2 底部给出的两个近似式

$$w_p = \tfrac{1}{2}(z_p + \sqrt{2k-1})^2 \tag{10}$$

或

$$w_p = k\left(1 - \frac{2}{9k} + z_p\sqrt{\frac{2}{9k}}\right)^3 \tag{11}$$

62 精确.

习题

1. 假设 Z 是标准正态随机变量，求：
   (a) $P(Z \leqslant 0)$.　(b) $P(Z \leqslant 1.96)$.　(c) $P(Z > 1)$.
   (d) $P(-1 < Z < 1)$.　(e) $P(-4 < Z < 0)$.　(f) Z 的上四分位数.
2. 假设 X 是具有均值为 0.5，标准差为 3 的正态随机变量. 求：
   (a) $P(X \leqslant 0)$.　(b) $P(X \leqslant 1)$.　(c) $P(X > -0.5)$.
   (d) $P(-1 < X < 1)$.　(e) X 的中位数.　(f) X 的上四分位数.
3. X 是某一高中运动员跑 1 英里所需要的时间（单位：分），假设 X 服从均值为 4.3，标准差为 0.05 的正态分布. 求该运动员在年度田径运动会上打破学校 4.15 分钟记录的概率是多少？
4. 令 X 是向一个大保险公司至少索赔过一次的保险客户的数量. 假设有 2000 位保险客户且每位客户全年索赔至少一次的概率为 0.2. 求任何给定的一年中，索赔的客户数不超过 500 的概率？
5. 如果某一班级中学生的体重近似于均值为 160、方差为 400 的正态分布，那么体重秤的最高刻度是多少才能使 99% 的学生能够称他们自己的体重？
6. 假设 Y 是参数 $n=60$，$p=0.5$ 的二项随机变量，试估计随机变量

   $$\frac{(Y-np)^2}{np(1-p)}$$

   超过 5 的概率.
7. 假设 W 是自由度为 k 的 χ^2 随机变量，求：
   (a) $k=4$ 时，W 的 0.95 分位数.
   (b) $k=8$ 时，W 的 0.95 分位数.
   (c) $k=200$ 时，W 的 0.95 分位数.
8. 假设 X, Y, Z 是独立的 χ^2 随机变量，自由度分别为 3、2、3. 求 W 超过 15 的概率，其中 $W = X + Y + Z$.
9. 设 X 是在为学校捐款活动中捐款的人数，假设有 500 人参加了这次活动，且每个人捐款的概率为 0.15，它们相互独立. 试估计 $P(80 < X)$.
10. 设 X 为十月份光顾得克萨斯州 Plains 市的 DQ 冰琪淋店至少 1 次的人数，假设 Plains 市有

2000居民，且每个人去DQ店的概率为0.25，彼此独立. 试估计$P(460<X<540)$. 63

11. 一名篮球运动员投篮100次有43次命中，如果此运动员真正的投篮命中率是60%，试求出他投篮命中不超过43次的概率.
12. 10个公司根据利润从1（最多利润）到10（最少利润）排名. 从这10个公司中随机取出4个，假设X是所选4个公司的排名和：
   (a) 如果$F(x)$是X的分布函数，求$F(14)$.
   (b) 用正态逼近求$F(14)$的近似值.

思考题

1. 假设W是自由度为100的χ^2随机变量，试用近似式（9），（10），（11）求得W的0.95分位数近似值，并与由表A2查得的精确值做比较.
2. 假设X是参数$n=100$，$p=0.3$的二项随机变量，利用表A1和A2，估计

$$P(20\leqslant X\leqslant 40)$$

1.6 第1章复习题

1. 如果得克萨斯理工大学所有录取的研究生中有75%是在得克萨斯理工大学的研究生招生计划中，并且所有申请学生中有40%的学生被录取，试求申请学生中被得克萨斯理工大学录取但属计划外的学生的百分比.
2. 投掷一枚均匀硬币8次，如果已知8次投掷中至少有两次是正面朝上的，试求恰有两次正面朝上的概率.
3. 投掷5个骰子，设X为这5个骰子上数字（1到6）的和. 假设骰子是均匀的（即每个骰子每次投掷等概率显示1到6），且彼此独立，求X的均值和方差.
4. 二维随机变量(X,Y)取$(0,0)$的概率是1/4，取$(1,1)$的概率是1/2，取$(2,0)$的概率是1/4，
   (a) 求X和Y的协方差. (b) X和Y独立吗？请解释.
5. 某俱乐部有8个成员，从成员中选取经理、副经理、秘书和会计的方式有多少种？
6. 小吃新品牌Yummies与品牌A和品牌B在进行投标竞争. 给4个宴会提供了等量的3种小吃，然后比较4个宴会后的剩余量. 假设Y是品牌Yummies被评为最受欢迎小吃的次数，如果在偏好上没有差别，则Y服从参数是$n=4$，$p=1/3$的二项分布. 64
   (a) 画出Y的分布图. (b) 求Y的下四分位数.
   (c) 求Y的四分位极差. (d) 求Y的均值.
   (e) 求Y的标准差. (f) 求Y不大于2的精确概率.
   (g) 运用正态逼近来估计Y不大于2的概率. 并与（f）中的值做比较.
7. 顾客从6个产品商标中等可能地选择任一个，标记“商标1”，“商标2”，依此类推. 假设X是选择的商标号. 若选择的商标号是前三个中的一个，则令$Y=3$；若是后三个中的一个，则令$Y=6$. 若选择的商标号为偶数，则令$Z=1$；若为奇数，则令$Z=2$.
   (a) 列出样本空间的点. (b) 写出样本空间上的概率函数.

(c) X服从什么分布？ (d) 求Y的四分位极差.

(e) 求Z的方差. (f) 求X和Z的协方差.

(g) Y和Z独立吗？

8. 12颗钻石根据质量从1到12排序．现从12颗中无放回地随机取出3颗，假设X是3颗钻石的序号的和.

(a) 样本空间中有多少点？ (b) 描述样本空间中的任一个点.

(c) 求$P(X=3)$. (d) 若$f(x)$是X的概率函数，求$f(10)$.

(e) 若$F(x)$是X的分布函数，求$F(10)$.

(f) 运用中心极限定理逼近$F(10)$，并将该近似值和（e）中求得的精确值进行比较.

9. 某高中毕业班前10名学生从1（最好）到10（第10名）排名．假设每一个名次等可能排上男学生和女学生，且X等于女学生名次的和．若前10名都是女生，则$X=1+2+3+\cdots+10=55$.

(a) 样本空间中有多少点？ (b) 描述样本空间中的任一个点.

(c) 描述样本空间上的概率函数. (d) 求$P(X=0)$.

(e) 求$P(X=1)$. (f) 若$f(x)$是X的概率函数，求$f(3)$.

(g) 若$F(x)$是X的分布函数，求$F(3)$.

10. 出生的婴儿中大约49%是女孩，51%是男孩．设一个家庭中有5个孩子.

(a) 女孩数的期望是多少？ (b) 女孩数的中位数是多少？

65 (c) 4个男孩和1个女孩的概率是多少？ (d) 男孩和女孩最可能的分布是什么？

(e) 为了回答这些问题，你还需要做哪些假设？

11. 两次独立地投掷一枚均匀骰子．假设$\bar{X}$是两次点数的平均数［即$\bar{X}=(X_1+X_2)/2$，其中X_1和X_2分别是第1次和第2次投得的点数］.

(a) 求出$\bar{X}=2$的概率. (b) 画出$\bar{X}$的整个概率分布条形图.

(c) 画出$\bar{X}$的分布函数图形. (d) 求$\bar{X}$的均值和方差.

(e) 求$F(3)$的精确值.

(f) 运用中心极限定理求$F(3)$的近似值，并与（e）中的值进行比较.

12. 假设预定某航班的人有10%并没有乘坐此次航班，正因为如此，航空公司为了乘载尽量多的人，通常会提供比飞机定员更多的预定．如果飞机能容纳100人，试问：航空公司能提供多少预定，又有90%的把握保证前来的每个预定的乘客能搭乘此次航班？

13. “Texas彩票”的玩法是，彩民从1到50数字中，无放回地（即数字不会重复出现）取出6个数字，然后彩票公司也从1到50中无放回地随机抽取6个数（每6个数字的组合等概率出现）．彩民若至少有3个数字和彩票公司取出的一样（不考虑数字顺序），则认为中奖.

假设X是彩民取出的匹配数字的个数，如果$X=3,4,5$或6时，则认为他中奖.

(a) 50个数字中选取6个数字的组合有多少（不考虑数字取出的顺序）？若数字随机取出，则每一组合出现的概率是多少？

(b) 求彩民取出的6个数字和彩票公司随机取出的数字完全匹配的概率，即$P(X=6)$是多少？

(c) 求$P(X=5)$,$P(X=4)$,$P(X=3)$各是多少？

(d) X 的概率分布是什么？概率分布中的参数值是多少？

(e) 在一次结果公布中，有 24 1024 张票都恰有 3 个数字正确，你认为这次彩票公司共卖出去多少张彩票？

(f) 在 (e) 的抽取中有 12 422 张票恰有 4 个数字正确，这与恰有 3 个数字相同的票数相容吗？

14. 参加“让我们做交易”游戏秀的竞猜者有机会获得一辆新轿车. 她所需要做的就是从 3 个一样的车库门（标号 A,B,C）中选择一个正确的车库门，该门后面藏着汽车. 她选择了 A.

在开车库门 A 之前，主持人 Monte Hall 问她是否需要改变主意. 为了让游戏更精彩，Monte Hall（知道车在哪）故意打开门 B，每个人都发现车不在里面，然后又问她是否需改变主意选门 C 而不是 A，此时，竞猜者应改变主意选 C 吗？ 66

(a) Monte Hall 开车库门 B 之前，门 A 是正确（即竞猜者获得那辆车）的概率是多少？

(b) Monte Hall 开车库门 B 之后，门 A 是正确的概率是多少？门 B 和门 C 是正确的概率又分别是多少呢？ 67

第2章　统 计 推 断

导　　言

前一章所介绍的概率论中的概念并没有涉及整个概率论领域，但这些简洁的介绍，对于帮助我们理解大多数常用的非参数统计方法中的基本原则则是必需的. 现在我们要架起概率论与其在数据分析中的应用之间的桥梁. 本章将要介绍数据分析的基础学科——统计（statistics）这一概念.

统计中很多的重要思想都要归功于在应用科学中处理数据时遇到困难问题的人们，他们具有一定的应用数学能力、一些数学训练及众多的常识. 他们的思想经过长期发展，浓缩为本章中我们逐步所要介绍的一些基本概念.

2.1　总体、样本与统计量

我们对所居住的这个世界的大多数认识都来源于样本. 我们在某家餐馆吃过一次饭，于是会对这家餐馆的饭菜质量和服务水平有一个看法. 我们结识了12个英国人，于是感觉自己差不多对所有英国人都有了一定的认识. 大多数情况下，从样本中获取的认识并不准确，但是，运用科学方法获得的样本却能够提供关于整个总体
68 的比较准确的信息.

试验

科学观点的形成常常源于试验（experiment）的框架. 正如我们在第1章所讨论的，一个试验就是每一个步骤都规定得很明确的过程，而在试验之前，每一步的结果都是未知的.

检验一个新药物治疗效果的试验由以下几部分组成：选定治疗病人，按照规定的步骤服用药物，观察该治疗方案的效果. 检验人工产品质量的试验则包括两部分：根据明确规定的步骤抽取和检验产品样本，记录试验结果.

总体

研究对象的全体所构成的集合称为总体（population），总体可以是一组人，一群动物，或甚至是诸如来自工厂产品装配线的零件之类的无生命物体. 一些总体比较小，例如美国历届总统组成的总体，这种情况可以考查整个总体. 有

一些总体比较大，像得克萨斯州的人口数；或基本上是无穷的，比如所有人类总人口，这种情况下对于总体的任何研究都只能基于从中抽取的样本.

样本

总体中某些元素的集合称为样本（sample）. 根据不同的获取方法，样本分为下面几个不同类型，方便样本（convenience sample）是一些最容易获得元素的集合，例如在街上采访的市民，或电视电话调查. 我们不太可能从这种样本中获得总体参数的精确估计. 另一方面，概率样本（probability sample）则能够相对精确地描述总体的未知参数，概率样本要求总体中每一个元素都有已知的非零概率. 本书中所考虑的概率样本是随机样本（random sample），这个概念我们将在本节的后面定义.

目标总体与样本总体

假如一名心理学家想要研究不停地打断一个人的睡眠对他情绪稳定的影响，他所考虑的总体应是当代的所有人. 为了进行试验，他在大学校报上刊登广告来招聘所需要的有偿志愿者. 他所抽取的样本很难具有代表性，因为这些志愿者都 69
是大学生，来自同一所大学，年龄范围相当狭窄，并且有某种相似的性情促使他们回应报纸上的广告，并应聘成为某项人体试验研究的志愿者. 但是，由于很多实际原因，比如有限的研究基金和时间，他不得不使用这种类型的样本，否则就得放弃整个试验. 因此有两种总体是值得一提的：研究的目标总体和实际样本的总体.

我们需要从中获取信息的总体称为目标总体（target population），而从中抽样的总体成为样本总体（sample population）. 上面的例子中考虑当代人类的全体作为目标总体，而来应聘的志愿者是样本总体. 所有的试验者都只能基于样本总体来研究问题，而试验的有效性取决于样本总体与目标总体相似的假设，至少在我们所研究的性质上是相似的.

随机样本

本书所讨论的统计方法通常假设样本是随机样本，所以介绍随机样本的有关概念是很重要的.

我们有两种方式来定义随机样本，第一种定义是总体元素的个数是有限的 N，这里 N 可以很大（世界上所有人口数）也可以很小（美国历届总统数）. 总体中每个元素的重要性相同，且等可能被抽取到. 容量为 n（$n<N$）的一组样本可以这样抽取：将总体中所有元素从 1 到 N 进行编号，从中随机抽取 n 个号码，使得出现任意 n 个号码的组合等可能，这 n 个号码对应着总体中的 n 个元素. 这种抽样方法通常是无放回（without replacement）的，所有相同的元素不会在样本中出现多于一次. 而对于有放回（with replacement）抽样的定义，相同的元素则可能出现两次或两次以上.

定义 2.1.1　从有限总体中任意抽取一组容量为 n 的样本，如果每组样本出现的可能性相等，那么称这样得到的样本为*随机样本*（random sample）.

上面定义中的“随机”不是针对样本本身，而是指获取样本的抽样方法，这一点看起来似乎有些奇怪. 事实上，我们是看抽样方法，而不是看样本本身来判断一组样本到底是不是随机样本.

假如一个有限总体共有 N 个元素，那么正如在 1.1 节所述，无放回抽样得到的
70 容量为 n 的样本共有 $\binom{N}{n}$ 种可能，有放回抽样样本共有 N^n 种可能. 若每组样本出现的可能性相等，则认为这样的抽样方法是随机的，得到的样本是随机样本.

当总体有限时，前面对随机样本的定义在大多数情况下是合适的. 但是，假如我们要考察某指定的人在一个晚上做梦的个数，可能会遇到麻烦. 在这种情况下，我们认为“随机样本”指某一晚做梦的个数，另一晚做梦的个数，直至比如说 7 个晚上做梦的个数. 即使在理想的情形下，这种抽样方法也不能符合定义 2.1.1 中的“等可能性”这一概念的框架. 什么叫等可能性？不是针对个体，因为前面我们假设的研究对象只是个体，不是总体的一个代表（尽管这可能是我们想要研究的最终目标）. 我们为了保证等可能性，难道要在这个人被期望能够活着的夜晚中，选择一些夜晚来做研究吗？显然，这是不可能的. 所以，随机样本至少还需要一个其他的定义.

数理统计中随机样本的标准定义如下所述：

定义 2.1.2　*容量为 n 的随机样本*（random sample of size n）是指一组 n 个独立同分布的随机变量列 $X_1, X_2, \cdots, X_n$.

在定义 2.1.1 中，如果抽样方法是*有放回*时，则定义 2.1.1 和定义 2.1.2 是相同的，并且当且仅当在这种情形下才是独立的. 无放回抽样产生的观测是非独立的，因为某个个体一旦被选中且不放回，就意味着它不可能再被抽取到. 然而，如果总体容量 N 很大，有放回抽样和无放回抽样在实际应用中的差别非常小，所以可以忽略这种观测间轻微的不独立性. 本书中的定理和公式的推导都假设样本中的观测是独立的. 对于有限总体，这些定理在其他假设下的修正是存在的，但不在本书的考虑范围之内. 这种修正的效果只要在样本量 n 小于总体容量 10% 的情况下就可以被忽略.

多元随机变量

试验者可能会测量或观测到定义 2.1.1 中随机样本的每个被选元素，以及定义 2.1.2 中的每个随机变量 X_i 的几个互相关联的特征，在这种情况下，用来描述几个特征的随机变量通常有两个脚标，比如 Y_{ij}，这里第一个脚标表示所选样本的个体，第二个脚标表示被测量或观测的某个特征.

也就是说，X_i 实际表示的是 k 维随机变量（$Y_{i1}, Y_{i2}, \cdots, Y_{ik}$），$X_i$ 仍然是独立同分

布的，但是 X_i 中的每个随机变量 Y_{ij} 可以是独立的，也可以是非独立的，可以是同分布，也可以是不同的分布.

举一个例子来说，考虑刚才讨论的“梦”的试验. 随机变量 X_i 表示第 i 个观测夜晚做的梦的个数，像定义 1.3.11 中定义的一样，假设 X_i 是独立且同分布（意思是每个 X_i 都有相同的分布函数）有一定的合理性. 但是如果试验者每晚不仅记录梦的总数，还记录整个睡眠时间，我们分别用 Y_{i1},Y_{i2} 表示，这样每晚做梦的个数和睡眠时间可能是相关的变量，所以 Y_{i1},Y_{i2} 很可能不是独立的. 但是，每个晚上的睡觉模式彼此是独立的. 在数学上，这就意味着 $Y_{i1},Y_{i2},Y_{j1},Y_{j2}$ 的联合概率分布函数可以分解如下： 71

$$f(y_{i1},y_{i2},y_{j1},y_{j2})=f_1(y_{i1},y_{i2})f_2(y_{j1},y_{j2}) \tag{1}$$

这里 f_1 和 f_2 分别是（Y_{i1},Y_{i2}）和（Y_{j1},Y_{j2}）的联合概率函数. 假如连续两晚睡觉模式的联合概率分布不变，即 f_1 和 f_2 一样，那么我们可以说（Y_{i1},Y_{i2}）和（Y_{j1},Y_{j2}）有相同的分布. 为了更方便地表达这种关系，即随机向量之间要求独立同分布，而随机向量内部的随机变量不必独立同分布，我们可以用 Y_{i1},Y_{i2} 的联合来表示 X_i，这时称 X_i 为二维随机变量. X_i 的值实际上包括两个值，一个是 Y_{i1} 的值，一个是 Y_{i2} 的值. 这样，前面所述的可以概括为“随机变量 $\{X_i\}$ 是独立同分布的”.

类似地，我们还可以考虑每晚有 k 个测量，它们是 $Y_{i1},Y_{i2},\cdots,Y_{ik}$，用 X_i 来表示这 k 个随机变量，那么称 X_i 为 k 维随机变量（k-variate random variable），或是多维随机变量（multivariate random variable）. 从定义 1.3.11 的角度来讲，X_i 是独立的就意味着所有 $\{X_i\}$ 的联合概率分布可以分解成 n 个联合概率函数的乘积，并且每个都是 $Y_{i1},Y_{i2},\cdots,Y_{ik}$ 的联合概率函数. 同样地，X_i 同分布是指上面提到的联合概率函数是相同的函数.

现在我们有两种随机样本的定义，第一种定义仅仅适用于有限总体样本并且直接与样本空间联系在一起. 如果每一种可能的样本（容量为 n）表示成样本空间中的一点，且样本空间中每个点被选为样本的概率相等，那么这种抽样方法是随机的，且抽得的样本是随机样本. 上面的定义中，我们仅用到样本空间以及概率函数的概念，但是并没有明确或含蓄地提及随机变量这一概念.

例 2.1.1

一个心理学家希望选取 4 名研究对象来进行个体训练和考试. 他登出广告，有 20 个志愿者应聘. 他有几个方法从容量为 20 的样本总体中抽取一容量为 4 的样本. 72

他可能会选择最先来应聘的 4 名志愿者，他的选择会偏向于那些积极主动的志愿者，这可能就不是随机样本.

他可能严格按照定义 2.1.1 来考虑，选择容量为 4 的样本，有 $\binom{20}{4}=4845$ 种可

能，那么他可以用4845张同样的纸，每张纸上写4个名字，每次都是一个不同的组合，然后把它们放到篮子里，随机地抽取一张纸片，纸片上的4个人则被选中. 这样得到的是随机样本，但这种抽样方法是不现实的.

另外一种获取随机样本的方法是，把20个名字各写在20张纸上，然后以某种随机方式一个接一个地抽取4张纸，比如可以从装满这些纸的一个帽子中抽取. 这种抽样方法同样满足随机样本的定义，这个过程可以通过计算机编程来模拟. ■

随机样本的第二个定义直接与随机变量相关，而不涉及样本空间. 但是，由于随机变量是定义在一个样本空间上的函数，尽管我们没有直接引进样本空间这一概念，但是它隐含在实际背景中. 同样，正如1.3节所提到的，随机变量所有可能取值的全体构成了样本空间，有时，为了解决出现的统计问题，将近似样本空间的点列举出来是必要的. 实际上，如果所有可能的测量结果（随机变量假设的值）都是样本空间中的点，那么就不会产生什么混淆. 我们通常认为这些测量结果是数值，但是有时测量的数值很难清楚地表达出来. 所以，我们最好讨论各种不同类型的测量.

度量尺度

度量的类型通常被称为*度量尺度*（measurement scale），各种不同的出版物都详尽地讨论过，其中包括Stevens(1946)的一篇优秀论文. 我们将从"最弱"的度量尺度，即名义尺度开始，通过讨论次序尺度和区间尺度，最后到"最强"的刻度，即比率尺度.

名义尺度

度量的*名义尺度*（nominal scale）只是使用数字将性质或元素分成不同种类或范畴的一种方法. 分配到观测上的数字只是用作"名字"以便说明观测所在的种类或
73 范畴，因此叫做"名义尺度". 对掷硬币我们定义随机变量为：硬币正面朝上时，它为1，反面朝上时，它为0，这时使用了度量的名义尺度. 我们也可以适当地选择7.3和3.9来分别表示正面和反面，我们选择0和1主要是因为方便计算所掷硬币中正面朝上的总次数. 当把12个研究对象用1到12个数字任意标号时，这时使用了度量的名义尺度，号码的分配则是随机变量的一种形式. 当根据颜色将研究对象分类时，种类可以用1,2,3或蓝、黄、红或A、B、C来标记. 这些号码只是类别的名字，当然只要种类保持不变，也可以用其他未使用过的号码来代替.

次序尺度

度量的*次序尺度*（ordinal scale）用于存在诸如"更小"，"更大"，"相等"这些比较关系的度量中. 度量的这些具体数字只是用来从小到大有序地排列元素的一种工具，由于它能够根据度量的相应大小对元素进行排序，所以称为*次序尺度*. 如果其中一些元素彼此相等，我们称为结. 当一个人用数字1来表示3个品牌中最喜欢的一个，3表示最不喜欢的一个，2表示剩下的那个品牌，这时，她就是在使用度量的次序尺度，数字只是作为表达她喜欢程度的一种方便方式. 当然，她可以用任意三

个数字如16,20,75来代替1,2,3,只要这些数字的相关顺序能够表达出她相应的喜欢程度就行.

区间尺度

第3种尺度是度量的*区间尺度*（interval scale），在一般的度量中，不仅考虑度量的次序尺度，还会考虑到两个度量区间的大小来作为相关信息，即两个度量间差别（从减法的意义上来讲）的大小. 区间尺度涉及一种单位长度的概念，任意两个度量间的距离可以用一些单位长度的倍数来表达. 用来理解区间尺度这一概念最好的例子就是我们日常生活中的温度的表示法. 温度增加一个单位（度）定义为温度计中一定体积水银柱的变化量. 因此任意两个温度的差别可以用这个单位，或度来衡量. 温度的实际数值只是和一个任选为“零度”点的比较. 测量的区间尺度需要一个零点和一长度单位（只有后者没有前者是不行的），但是哪点定义为零点，哪种长度定义为单位长度并不重要. 温度可以同时由华氏温标和摄氏温标来计量，它们有不同 74 的零度和不同定义的1度或单位. 区间度量的法则不会因刻度或位置或两者同时的改变而受干扰.

比率尺度

最后，不仅当次序和区间的大小很重要，而且两度量的比率也很有意义时，我们需要引入度量的*比率尺度*（ratio scale）. 如果说一个量是另一个量的“2倍”是合理的话，引入度量的比率尺度就是合适的，如度量农作物产量，距离，重量，高度，收入等. 实际上，比率尺度和区间尺度的唯一差别是前者要求有绝对零点，而后者的零点可以是任意一点，和区间尺度一样，比率尺度的单位长度也是可以任意定义的.

我们不可能就度量本身来谈哪种度量尺度是合适的，而应该考虑被度量的量以及度量方法，然后再决定赋予度量数值的涵义.

关于这4种度量尺度，科学家们没有达成一致的意见. 有些科学家喜欢用其他尺度，而有些度量也不能清楚地归类于上面4种尺度的任何一种. 这样看来，上面的分类显得把问题过于简单化，但针对本书目的而言已经足够了.

大多数常用参数统计方法要求度量是区间尺度（或者比这更强的尺度），而大多数非参数统计方法通常假设名义尺度和次序尺度是合适的. 当然，每种度量尺度应有弱度量尺度的所有性质. 因此，只需要弱度量的统计方法可能也会用强度量.

统计量

到目前为止，我们已经讨论了总体，来自总体的样本，以及度量样本所感兴趣的性质的度量尺度. 度量尺度涉及随机变量，因为度量样本元素的体系实际上就是一个随机变量. 由于统计量是随机变量，因此，度量尺度与统计量有关. 对于数理统计学家来说，“统计量”和“随机变量”这两个术语是可以互换的. 但是，统计量

一词的普遍使用表明它不仅仅是一个随机变量.

统计量一词本来是指国家公布的由政府收集的数据总括的结果，因此有人认为
75 统计量就是基于一些数的一个数，比如样本均值，总体中某一类元素占整个总体的比例等等. 从这个意义上讲，统计量就是一个数. 但是，如果我们考虑不同样本均值具体数值存在不同，或是不同时间总体富有变化，我们就能够将统计量的概念从仅仅一个数扩展到得到这个数的法则. 这时，“样本均值”就是统计量. 一个样本中实际得到的平均值就是统计量的一个值，作为法则，统计量需要满足作为一个随机变量，样本空间（合理定义的样本空间）中点的函数的要求. 统计量还要体现数据总括这一想法，因此通常所考虑的统计量是几个随机变量函数的随机变量，统计量的值是这几个随机变量值经过算术运算所得的结果. 由于随机变量是定义在样本空间的函数，那么统计量则是定义在一个特殊样本空间上的函数，这个样本空间中的样本点是 n 维随机向量的所有可能值. 下面统计量的正式定义和例子将进一步阐明这个概念.

定义 2.1.3　一个统计量（statistic）是将样本空间中的样本点映射到实数上的函数，其中样本空间中的样本点是一些多元随机变量的所有可能值. 换句话说，统计量就是几个随机变量的函数.

作为统计量的定义，定义 2.1.3 中的每一句话都是充分的，它们清楚地阐述了这个概念.

例 2.1.2

用 $X_1,X_2,\cdots,X_n$ 表示 n 个学生的考试分数，每个 X_i 都是随机变量. 令 W 等于考试分数的平均值，

$$W=\frac{X_1+X_2+\cdots+X_n}{n}=\frac{1}{n}\sum_{i=1}^{n}X_i \tag{2}$$

则 W 是一统计量. 若 $X_1=76,X_2=84,X_3=85$ 表示 3 个学生的考试分数，$W=(\frac{1}{3})(76+84+85)=81\frac{2}{3}$. 统计量 W 满足定义 2.1.3 中的第二句话：它是随机变量 $X_1,X_2,\cdots,X_n$ 的函数. 由于 W 将随机向量（$X_1,X_2,\cdots,X_n$）映射到实数，这满足定义 2.1.3 中的第一句话. 这时，若多元随机向量（X_1,X_2,X_3）的值为（76,84,85），那么统计量 W 的值为 $81\frac{2}{3}$. 统计学中经常应用这一特殊的统计量，称为“样
76 本均值”，下一节中将进一步讨论它. ■

次序统计量

我们经常会使用称为*次序统计量*（order statistic）的一类特殊的统计量，特别当我们要处理有序测量数据的时候. 假如一随机变量（$X_1,X_2,\cdots,X_n$）的观测值（$x_1,x_2,\cdots,x_n$）是“有序的”，即它的元素是按从小到大的顺序排列的，我们用 $x^{(1)}\leqslant x^{(2)}\leqslant\cdots\leqslant x^{(n)}$ 来表示*有序观测*（order observation）.

定义 2.1.4 把 $(X_1,X_2,\cdots,X_n)$ 的每个观测值 $(x_1,x_2,\cdots,x_n)$ 按从小到大排列，取值为第 k 个值 $x^{(k)}$ 的随机变量成为*秩为 k 的次序统计量*（order statistic of rand k）$X^{(k)}$.

因此，秩为 1 的次序统计量 $X^{(1)}$ 总是取 $(x_1,x_2,\cdots,x_n)$ 中最小值. 在例 2.1.2 中，$X^{(1)}=76, X^{(2)}=84, X^{(3)}=85$. 若 (X_1,X_2,X_3) 的另一观测取值 (93,73,81)，那么次序统计量为 $X^{(1)}=73, X^{(2)}=81, X^{(3)}=93$. 若 $(X_1,X_2,\cdots,X_n)$ 是一随机样本，则有时 $X^{(1)}\leqslant X^{(2)}\leqslant\cdots\leqslant X^{(n)}$ 称为*次序随机样本*（ordered random sample）.

我们将会在 2.2 节中介绍很多其他有用的统计量，并进一步讨论这些统计量在分析试验结果中的作用.

习题

1. 美国国会委员会希望检验关于美国高中立法的作用，从华盛顿地区随机选取 5 所中学来进行研究.
   (a) 目标总体是什么？
   (b) 样本总体是什么？
   (c) 如果华盛顿地区有 100 所高中，那么共有多少个不同的样本？
   (d) 在问题 c 中每个样本出现的概率各是多少？
2. 托皮卡一家电视台向观众提出这样一个问题："在堪萨斯州可否允许饮酒？" 372 个电话回访中有 164 人说"不可"，其余的人为"可以".
   (a) 目标总体是什么？
   (b) 样本总体是什么？
   (c) 样本是随机样本吗？请说明原因.
   (d) 题目中隐含了 3 个统计量，它们分别是什么，观测值取多少？
   (e) 统计投票数时采用了哪种度量尺度？
   (f) 记录电话回访的答案为"不可"或"可"时采用了哪种度量尺度？ 77
3. 田径运动会授予在比赛中积分最高的运动队奖品，队中的运动员在比赛中每次获得第一、第二、第三名的积分值分别为 5,3,1.
   (a) 记录积分采用了哪种度量尺度？
   (b) 题目中提到的（隐含）统计量是什么，有何作用？
4. 足球队员队服上印有不同的号码，这些号码采用的是哪种度量尺度？
5. 下面采用的是哪种度量尺度？
   (a) 邮政编码　(b) 本地电话号码
   (c) 电话区号　(d) 社会保险号码
6. 下面采用的是哪种度量尺度？
   (a) 月工资　(b) 汽油泵上被度量的加仑数
   (c) 每磅咖啡的价格　(d) IQ 分数表示的智商
7. 为了随机选取一律师事务所，从这个城市中所有律师的名单中随机抽取一个，则这个律师所在的律师事务所则被选中. 问这个律师事务所的选取是随机的吗？

8. 我们采用下面的方法来估计观看各种电视节目观众的数量：2200 个家庭作为随机样本进入调查，这些家庭同意将他们的电视机和一电子设备连接，以便能追踪他们所观看不少于 8 分钟的节目.
   (a) 目标总体是什么？　(b) 样本总体是什么？
   (c) 评述结论的精确度.

思考题

1. 一个试验的研究对象是掷 n 次不均匀型骰子. 令 X_i 表示第 i 次投掷的骰子点数，那么 $X_1, X_2, \cdots, X_n$ 组成的是一组随机样本吗？
2. 从整数 1 到 7 中无放回地抽取容量为 4 的一组随机样本，
   (a) 可能样本的总数共有多少？
   (b) 每个样本的概率是多少？
   (c) 样本中至少有一个奇数的概率是多少？
78 (d) 样本中数的总和为 12 的概率是多少？
3. 从整数 1 到 N 中无放回地抽取容量为 n 的一组随机样本，样本中至少有一个奇数的概率是多少？

2.2　估　　计

统计量的一个基本目的是估计总体的未知性质. 这些估计出的未知性质通常是用数字表示的，并且包括可列举的一些项目，例如未知比率、均值、概率等等. 事实上，估计是基于样本（如果有概率描述，则是随机样本）的，并且估计是关于随机变量分布未知性质的有根据推测，这里随机变量表示对总体研究感兴趣的量. 例如，我们可以用晶体管产品中样本的不合格率来估计总体的不合格率. 用来作估计的统计量自然叫做*估计量*（estimator）. 本节我们将要讨论一些估计量，例如*样本均值*（sample mean），*样本方差*（sample variance）和*样本分位数*（sample quantiles）. 我们首先引入一个与众不同的估计量，*经验分布函数*（empirical distribution function）.

经验分布函数

一个随机变量的真实分布函数一般是未知的，有时，我们只能够推测分布函数的形式，或将推测作为真实分布函数的一个近似. 根据样本的观测值构作 $S(x)$ 图，以此来作为整个未知分布函数 $F(x)$ 的估计，这是推测分布函数的一种好方法. 介绍作图方法最好用举例子来解释，由此我们给出定义：

定义 2.2.1　设 $X_1, X_2, \cdots, X_n$ 是一组随机样本. *经验分布函数* $S(x)$（简称为 e. d. f）是 x 的函数，它在 x 点的取值为小于或等于 x 的 X_i 在样本总数中所占的比例，$-\infty < x < \infty$.

例 2.2.1

在一项体能研究中，从某一高中随机抽取了 5 名男生，记录他们跑完 1 英里的时间. 时间（转化成分钟后）分别为 6.23, 5.58, 7.06, 6.42, 5.20，把它们标记在图 2-1 横轴上. 由于经验分布函数 $S(x)$ 是小于或等于 x 的 X_i 在样本总数中所占的比例，根据这组特定样本，把它画在图 2-1 中. 79

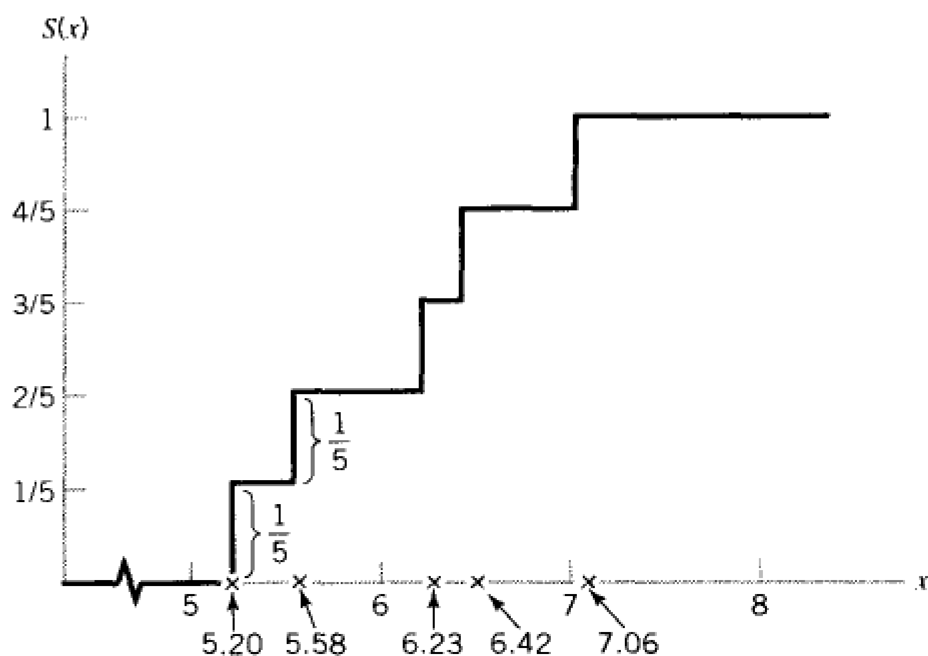


图 2-1 经验分布函数 ■

从例 2.2.1 中可以看出，经验分布函数总是阶梯函数，每阶的高度是 $1/n$，并且只在样本取值处有变化. 图 2-1 中竖线并不是经验分布函数的一部分，只是一方面为了外观，另一方面在后来确定样本分位数时比较方便. 我们从左到右来考虑经验分布函数的图像，注意到 $S(x)$ 在样本最小值前均取值为零，在每个样本取值处会增加一阶的跃度，每个跃度是 $1/n$. 在样本最大值处 $S(x)$ 取最大值 1.0，并且在剩下所有比样本最大值大的 x 处都取 1.0. $S(x)$ 很像非降、取值从 0 到 1 的分布函数. 但 $S(x)$ 只是由经验（来自样本）确定的，并由此而得名.

图 2-1 只描述了 $S(x)$ 的一组观测值，其他的样本值将产生另外不同的 $S(x)$ 的图像. 这表明了 $S(x)$ 的随机性，从这个意义上讲，它是一个随机变量. 但是，由于它是一个函数，且观测值是整个图像而不是单个值，所以称 $S(x)$ 为**随机函数**（random function）更加合适. 因为它能够相当好地估计随机变量的分布函数，所以它通常用做一个估计量. 为了区分经验（或样本）分布函数，我们称随机变量的分布函数为总体分布函数.

从某种意义上讲，经验分布函数的观测值可以认为是总体分布函数的取值，准确些讲，基于样本观测值 $x_1, x_2, \cdots, x_n$, $S(x)$ 的一个观测值和一个取 $x_1, x_2, \cdots, x_n$ 中每个值概率都是 $1/n$ 的随机变量的分布是一样的. 这种随机变量的分布函数是一个阶梯函数，且在每个数值 $x_1, x_2, \cdots, x_n$ 处的跃度为 $1/n$. 利用第 1 章中的定义，我们容易得到随机变量的均值，方差和分位数. 80

例 2.2.2

随机变量 X 的分布函数与例 2.2.1 中 $S(x)$ 相同，它有如下概率分布：

$$P(X=5.20)=0.2 \qquad P(X=5.58)=0.2$$
$$P(X=6.23)=0.2 \qquad P(X=6.42)=0.2$$
$$P(X=7.06)=0.2$$

X 的分布函数的图像与图 2-1 相同. 由定义 1.4.1 知，X 的中位数是 6.23. 由定义 1.4.3 知，X 的均值为

$$\begin{aligned} E(X) &= \sum_x xf(x) \\ &= (5.20)(0.2)+(5.58)(0.2)+(6.23)(0.2)+(6.42)(0.2)+(7.06)(0.2) \\ &= 6.098 \end{aligned} \tag{1}$$

同样，由定义 1.4.4，可计算 X 的方差为

$$\begin{aligned} \mathrm{Var}(X) &= \sum_x (x-E(X))^2 f(x) \\ &= 0.424 \end{aligned} \tag{2}$$

■

估计量

为了区分真实"总体"的均值，方差和分位数，由样本计算得到的均值，方差和分位数（如在例 2.2.2 中）分别称为*样本均值*，*样本方差*，*样本分位数*. 正如经验分布函数可以作为总体分布函数的估计量，样本均值，方差，分位数也可以分别作为总体均值，方差，分位数的估计量.

81 **定义 2.2.2** 设 $X_1,X_2,\cdots,X_n$ 是一组随机样本，样本 p 分位数 Q_p 满足以下两个条件：

1. 小于 Q_p 的 X_i 的比例 $\leq p$.
2. 大于 Q_p 的 X_i 的比例 $\leq 1-p$.

正如总体分位数从总体分布函数得到的方式一样，每个样本分位数都可以由经验分布函数得到. 样本 p 分位数是 $S(x)=p$ 处的 x 值，如果不止有一个 x 值满足 $S(x)=p$，我们取最大值与最小值的均值作为该样本 p 分位数，与总体分位数的处理一样. 样本 p 分位数 Q_p 取决于随机变量的取值，因此它是一个统计量. 注意，为简便起见，我们只针对随机样本定义样本分位数.

一种直接由样本而不通过 $S(x)$ 的图像而得到样本 p 分位数的方法是，用 p 乘以样本容量 n，四舍五入得到一相邻的较大整数，以该整数为秩的次序统计量的观测值就是样本 p 分位数. 如果 $(p\cdot n)$ 是一整数，那么样本分位数是以 $(p\cdot n)$，$(p\cdot n+1)$ 为秩的两个次序统计量的平均值.

例 2.2.3

从市区妇女俱乐部中的已婚妇女中随机抽取 6 名妇女，记录每位妇女的孩子个数，分别为 0,2,1,2,3,4，经验分布函数如图 2-2 所示. 样本中位数 $Q_{0.5}$ 是 2，样本分位数 $Q_{0.25}$，$Q_{0.75}$ 分别为 1 和 3. 按照我们的约定，1/3 样本分位数 $Q_{1/3}$ 是 1

和 2 的平均值，即 1.5. 这些数可以用来估计未知的总体分位数.

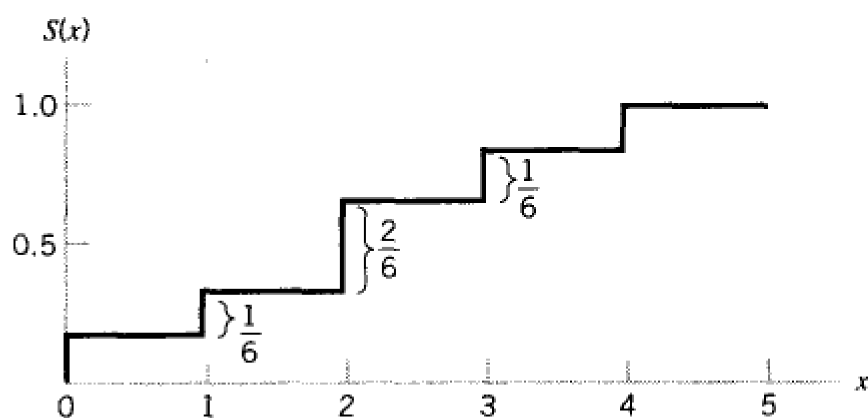


图 2-2 经验分布函数 ■

注意，在（1）和（2）式中，$f(x)=1/n$，可以把这一因子提到和号外，这样样本均值和方差计算起来比例 2.2.2 简单. 由这种简便的计算方法，我们给出下面的定义. 82

定义 2.2.3 设 $X_1, X_2, \cdots, X_n$ 是一组随机样本，则*样本均值* $\overline{X}$ 定义如下，

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \tag{3}$$

样本方差 S^2 定义如下，

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 \tag{4}$$

同时，它还等价于下式

$$S^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2 \tag{5}$$

样本标准差（sample standard deviation）S 是样本方差的平方根.

例 2.2.4

例 2.2.3 中随机样本 0,2,1,2,3,4 的样本均值是

$$\begin{aligned}\overline{X}&=\tfrac{1}{6}(0+2+1+2+3+4)\\&=2\end{aligned} \tag{6}$$

样本方差是

$$\begin{aligned}S^2&=\tfrac{1}{6}(2^2+0+1^2+0+1^2+2^2)\\&=1\tfrac{2}{3}\end{aligned} \tag{7}$$

因此，未知均值的估计是 2，未知方差的估计是 $1\frac{2}{3}$. ■

可能除了经验分布函数外，上面所介绍的估计量提供了未知总体参数的*点估计*（point estimate）. 即，前面的例子中我们得到了未知均值的估计，“均值的估计值为 2”，单点“2”是估计.

我们经常更喜欢，同时也更谨慎地说：“我们以 95% 的置信水平认为未知均值落在 1.3 与 2.7 之间.” 这种估计称为*区间估计*（interval estimate）. 区间估计量由两个

统计量组成，它们是区间的两个端点. 置信系数（confidence coeffcient）是区间估计量包含未知总体参数的概率. 前面的叙述中置信系数是0.95. 区间和置信系数一起
83 称为未知量的置信区间.

点估计是比较容易的，因为作点估计只需要考虑一个数，任意一个数. 但是，有些点估计量比其余的估计量要好. 为了比较哪种估计量更好，其点估计的比较标准几乎在任何一本概率统计导论的书中都可以找到.

一个好估计量的标准之一是无偏性（unbiased）. 在下面的讨论中，我们通常用希腊字母 θ,μ,σ 或 ρ 表示参数，$\hat{\theta}$ 表示用来估计 θ 的统计量.

定义 2.2.4 如果 $E(\hat{\theta})=\theta$，则称统计量 $\hat{\theta}$ 是总体参数 θ 的无偏估计（unbiased estimator）.

下面的定理表明 $\bar{X}$ 是总体均值的无偏估计.

定理 2.2.1 $X_1,X_2,\cdots,X_n$ 是一组来自均值为 μ，方差是 σ^2 总体的独立随机变量，那么

$$E(\overline{X})=\mu \tag{8}$$

并且

$$\mathrm{Var}(\overline{X})=\sigma^2/n \tag{9}$$

证明 由定理1.4.1，

$$E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)=n\mu \tag{10}$$

所以

$$E(\overline{X})=E\left(\frac{1}{n}\sum X_i\right)=\frac{1}{n}\sum E(X_i)=\frac{1}{n}n\mu=\mu \tag{11}$$

这表明 $\bar{X}$ 是总体均值的无偏估计，同时，由定理1.4.3

$$\mathrm{Var}\left(\sum X_i\right)=\sum \mathrm{Var}(X_i)=n\sigma^2 \tag{12}$$

那么，经过简单的代数运算，可以得到下式

$$\mathrm{Var}(\overline{X})=\mathrm{Var}\left(\frac{1}{n}\sum X_i\right)=\frac{1}{n^2}\mathrm{Var}\left(\sum X_i\right)=\frac{1}{n^2}n\sigma^2=\sigma^2/n \tag{13}$$

84 即完成了定理的证明.

标准误差

一个估计量的标准差通常称为标准误差（standard error），所以它不会和总体标准差的概念混淆，因为这是一个完全不同的概念. 如定理2.2.1所示，$\bar{X}$ 的标准误差是 $\sigma/\sqrt{n}$.

无偏估计量 s^2

我们已经看到 S^2 不是 σ^2 的无偏估计，因此，习惯上我们使用无偏估计 s^2

$$s^2=\frac{1}{n-1}\sum(X_i-\overline{X})^2 \tag{14}$$

作为 σ^2 的估计量. 但是, S 和 s 都是总体标准差 σ 的有偏估计.

渐近置信区间

由定理 2.2.1 及中心极限定理（定理 1.5.2），如果 $X_1, X_2, \cdots, X_n$ 是一组均值为 μ，方差是 σ^2 的独立随机变量，那么，当 n 趋于无穷时，

$$\frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{15}$$

的分布函数趋于标准正态分布. 在实际应用中，如果 n 足够大，

$$P\left(-z_{1-\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) \tag{16}$$

的概率近似是 $1-\alpha$，其中 $z_{1-\alpha/2}$ 表示标准正态随机变量的 $(1-\alpha/2)$ 分位数.

上面的不等式经过代数整理，可写为

$$P\left(\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \cong 1-\alpha \tag{17}$$

当 n 很大时，上式给出了 μ 的近似置信区间. 进一步讲，既然 σ 很少已知，那么通常情况下，当 n 足够大，对于来自有非零方差有限总体的随机样本，由中心极限定理我们可以用 s 来估计 σ，从而得到 μ 的近似置信区间. 对大多数的实际问题来说，当样本容量超过 30 就可以认为是“足够大”了. 85

例 2.2.5

一窝猪的数量越多意味着农场主可获得的利润就越多. 国家试验中心正在研究一种能够提高每窝猪产量的新技术，记录的 55 窝猪中平均每窝存活猪的数量是 9.8，且 $s = 1.4$. 这些平均每窝存活猪的总体均值 95% 的近似置信区间下限为

$$\overline{X} - z_{1-\alpha/2}\frac{s}{\sqrt{n}} = 9.8 - 1.96\frac{1.4}{\sqrt{55}} = 9.43 \tag{18}$$

近似置信区间上限为

$$\overline{X} + z_{1-\alpha/2}\frac{s}{\sqrt{n}} = 9.8 + 1.96\frac{1.4}{\sqrt{55}} = 10.17 \tag{19}$$

从而，可以以 95% 的置信水平认为真实的每窝猪数量的均值在 9.43 与 10.17 之间. ■

自助法

定理 2.2.1 给出了 $\bar{X}$ 的均值和方差，S^2 的均值也不难得到. 但是，S^2 的方差则不是那么容易计算. 很多统计量作为总体参数的估计量像定理 2.2.1 那样从理论上来推导是非常困难，甚至是不可能的，所以我们用其他的方法来估计它们的均值和方差. 其中一种方法称为*自助法*（bootstrap）.

自助法是从原始容量为 n 的随机样本的观测值中有放回地抽取 n 个值，也就是说，一些原始随机样本的观测值在“自助样本”中可能出现一次，多于一次，或者根本不出现. 自助样本的个数总是和原始随机样本中观测值的个数相等.

每组自助样本都可以计算出我们所关心的估计量 $\hat{\theta}$. 通过计算机模拟，我们可以由原始随机样本的观测值得到成百上千组自助样本，每组自助样本都会产生 $\hat{\theta}$ 的一个值. 这成百上千个 $\hat{\theta}$ 估计值的样本均值，样本标准差（s 或 S）可以用来估计 $\hat{\theta}$ 的总体均值和总体标准差（标准误差）. 事实上，在自助法中，这些 $\hat{\theta}$ 估计值的经验分布函数可以作为 $\hat{\theta}$ 的真实的总体分布函数的一个估计.

显然，自助法每一步都依赖于原始随机样本值，不同样本值的集合会产生不同
86 估计值的集合.

自助法重复试验次数

对于简单估计一个估计量的均值和标准差，自助法的重复次数很少超过 100 或 200，25 次左右已经足够. 但是，要得到置信区间则需要做大量的重复试验. 得到 θ 近似置信区间的一种方法是利用自助样本估计量 θ^* 的 $\alpha/2$ 和 $1-\alpha/2$ 的样本分位数. Efron 和 Tibshirane(1986)建议至少要做 250 次自助重复试验，他们同时也给出了另外一种更加精确获得置信区间的方法，这种方法需要更多的自助重复试验，至少要做 1000 次.

例 2.2.6

在例 2.2.5 中，对 $\bar{X}$ 用中心极限定理，得到了每窝猪数量总体均值的置信水平为 95% 的近似置信区间. 现在我们用自助法，求出每窝猪数量总体标准差 σ 的置信水平为 95% 的近似置信区间，这个参数对检验是非常有用的，因为每窝猪数量相差不大的情况（σ 较小）要比一些窝数量很少而其他的很多（σ 较大）好得多.

原始样本 55 个观测值是从 1 到 55 的编号.

观测号	1	2	3	4	. . .	55		
每窝数量	9	9	8	6	. . .	11	(55个)	$s=1.4$

现在从 1 到 55 号进行有放回地抽样，得到 55 个数，得到第一次自助样本，由它计算出估计量 s^*.

自助样本#1:

观测号	4	17	4	28	. . .	16		
每窝数量	6	9	6	10	. . .	9	(55个)	$s_1^*=1.6$

这个过程重复 250 次（自助法的这个过程可以重复所需要的尽可能多的次数，但是建议求置信区间至少需要 250 次).

自助样本#2:

观测号	28	23	3	1	. . .	39		
每窝数量	10	10	8	9	. . .	8	(55个)	$s_2^*=1.8$

将这个过程继续重复下去，直到

自助样本#250:

观测号	6	1	55	14	. . .	17		
87 **每窝数量**	10	9	11	11	. . .	9	(55个)	$s_{250}^*=1.1$

由 s^* 的样本 0.025 分位数可以得到 95% 置信区间的近似置信下限，因为

0.025(250) = 6.25，四舍五入到 7，所以置信区间的下限是第 7 个次序统计量，$s^{*(7)}$. 由样本 0.975 分位数，即第 244 个次序统计量（0.975 × 250 = 243.75，四舍五入到 244）得到，$S^{*(244)}$即为置信上限. 在这个问题中，我们需要将 s^* 的 250 个值从小到大排序，即

0.7, 0.8, 0.8, 0.9, 0.9, 0.9, 1.0, . . . , 2.0, 2.0, 2.2, 2.3, 2.3, 2.4, 2.7

从而，95% 置信区间是从 1.0 到 2.0. 通过计算 250 个 s^* 的标准差 $\hat{\delta}$ 而得到 s 的标准误差的估计，正如由 250 个 s^* 的均值 $\bar{X}$ 可以算出 s 均值的估计值一样，我们还可以算出其他所关心的统计量. ■

计算机辅助

几乎所有的计算机中的统计软件包，甚至许多便宜的手动计算器都可以计算出我们前面所讨论的点估计. 但是，自助法可不太容易得到. *S-Plus*, *SYSTAT*, *Resampling Stats* 和 *Stata* 等统计软件中都可以找到自助法的计算程序. 如要查看有关程序方面的具体细节，可参考 Davison 和 Hinkley(1997)写的关于再抽样的书中有关*S-Plus* 使用方法的介绍.

一般参数估计

用两个问题来概括估计一个未知参数 θ 的大致过程：

1. 使用哪个统计量？建议仿照我们前面所举的例子中计算 $\hat{\mu}=\bar{X}, \hat{\sigma}=S$ 以及分位数估计量 $\hat{x}_p = Q_p$ 的过程，看参数 θ 是怎样从总体分布函数 $F(x)$ 中定义的，就怎样从经验分布函数来定义参数 θ 的估计 $\hat{\theta}$.

2. 该统计量是否优良？这通常用一个估计量的标准差（称为*标准误差*）来衡量. 定理 2.2.1 给出了 $\bar{X}$ 的标准误差是 $\sigma/\sqrt{n}$. 其他统计量的标准误差并不像均值这样容易计算，但可以通过自助法来估计，自助法还可以给出 $\hat{\theta}$ 的整个分布函数的估计以及 θ 的近似置信区间.

□**理论** 这种估计方法的理论基础是随着 n 的增大，$S(x)$依概率趋于 $F(x)$（本书中没有提及"依概率"这个概念的精确定义），因此可以用 $S(x)$来估计 $F(x)$中的参数. 通常大多数情况下所关注的估计量都是趋于（依概率）被估计的未知参数，这使得它们成为一些较为优良的估计量. 从 $F(x)$中重复抽样可以得到估计量的渐近分布函数，所以只要样本容量足够大，从 $S(x)$中重复抽样得到的结果几乎与前者没有太大的差别. 关于自助法概念的介绍，请查阅 Efron 和 Tibshirani(1986)的参考文献. □ 88

生存函数

经验分布函数 $S(x)$将随机样本 $X_1, X_2, \cdots, X_n$ 与总体分布函数 $F(x)$有机地结合起来，因为它用小于等于 x 观测值的频率来估计 $P(X \leqslant x) = F(x)$. 在寿命测试，医疗后续工作，以及其他领域中，和分布函数同样有用的是*生存函数*(survival function)$P(x) = 1 - F(x)$，这里所关心的变量为事物的*寿命*（lifetime）（从发生到结束持续的时间），可以是人，动物或无生命产品的寿命，也可以简单地从某个起

点开始，直到某事件发生，如痊愈，到达，离开等等所持续的时间.

$P(x)$的一个自然估计是它的经验生存函数

$$\hat{P}(x) = 1 - S(x) \tag{20}$$

它是样本 $X_1, X_2, \cdots, X_n$ 中超过 x 值的频率.

Kaplan-Meier 估计

Kaplan 与 Meier(1985)用 X 表示死亡时间，应当注意，由于试验中某些项的缺失(loss)，死亡时间在某些情况下是不可观测的，比如在试验中研究对象的离开，进入试验研究较晚，或者试验结束后才死亡等等. 他们提供一种从缺失数据中获取信息的方法，即在缺失前死亡（death）还没有发生. 他们利用了下面的事实：如果死亡发生在时刻 x 后，那么在 x 前的任意时刻后死亡仍然会发生. 下面是他们的理由.

由条件概率的定义（等式 1.2.2），我们可以得到，对于 $x_0 < x_1$，

$$P(X > x_1) = P(X > x_1, X > x_0) = P(X > x_1 \mid X > x_0)P(X > x_0) \tag{21}$$

假设在第 1 年初有 100 个研究对象参加测试，第 1 年底只有 30 个存活，我们用下式估计 $P(1)$

$$\hat{P}(1) = \hat{P}(X > 1) = 30/100 = 0.3 \tag{22}$$

这里 X 表示研究对象个体的寿命.

接着，假设第 2 年初又有另外 1000 个个体参加试验，第 2 年底，1000 个中有 250 个存活，而最初 100 个中存活的 30 个只剩了 10 个. 我们可以用最初的 100 个个体来估计 $P(2)$，

89

$$\hat{P}(2) = \hat{P}(X > 2) = 10/100 = 0.1 \tag{23}$$

但此时，我们可以用第 2 年新参加的 1000 个个体的信息来更新估计 $P(1)$，因为一年中参加的试验的个体共有 1100 个，其中共有 $250 + 30 = 280$ 个存活，改进后 $P(1)$的估计为

$$\hat{P}(1) = \hat{P}(X > 1) = 280/1100 = 0.255 \tag{24}$$

进一步讲，由 1100 个个体中有 280 个存活和等式 2.2.21，我们可以用改进后的 $P(1)$ 来改进 $P(2)$，得到估计式

$$P(2) = P(X > 2) = P(X > 2 \mid X > 1)P(X > 1) \tag{25}$$

改进后的估计 $\hat{P}(x>1)$为 0.255. 不幸的是，我们无法改进 $P(X>2|X>1)$的估计值，因为我们不知道在接下来一年的试验中 1000 个观测有多少个存活. 所以我们用下面的估计量

$$\hat{P}(X > 2 \mid X > 1) = 10/30 \tag{26}$$

因为它仅用到了已知信息，即第 1 年底有 30 个存活，第 2 年底有 10 个存活，那么 $P(2)$的改进估计就是

$$\hat{P}(2) = \hat{P}(X > 2 \mid X > 1)\hat{P}(X > 1) = \frac{10}{30}\frac{280}{1100} = 0.085 \tag{27}$$

Kaplan 与 Meier 推广了上面的方法，应用到了一般的情况．设 $u_1<u_2<\cdots<u_k$ 表示 k 个个体"寿命"，这里寿命是指从开始到死亡，或者到研究缺失所持续的时间．并令

$$p_i = P(X > u_i \mid X > u_{i-1}) \tag{28}$$

用下式估计

$$\hat{p}_i = \frac{\text{到时刻 } u_i \text{ 的存活的个体数}}{\text{在时刻 } u_{i-1} \text{ 仍然观测到的存活个体数}} \tag{29}$$

在时刻 u_i 缺失的个体，可以认为在时刻 u_i 以后仍然存活，因为知道他们在时刻 u_i 还是活着的．而在时刻 u_i 死亡的个体不可能在时刻 u_i 后是存活的．在第 1 次死亡或缺失的计算时，$\hat{P}_1$ 的分母是参加试验个体的总数． 90

$P(x)$ 的 Kaplan-Meier 估计为

$$\begin{aligned}\hat{P}(x) &= 1 \quad \text{其中} x < u_1 \\ &= \prod_{u_i \leqslant x} \hat{p}_i \quad \text{其中} x \geqslant u_1\end{aligned} \tag{30}$$

其中，乘积是对所有寿命 $u_i \leqslant x$ 的 i 进行的．注意，这个估计量是一个递降的阶梯函数，且只在观测的死亡时间取值发生变化，而且由 $S(x)=1-\hat{P}(x)$，这种方法可定义更一般的经验分布函数 $S(x)$，用于针对缺失数据的研究．

计算机辅助

Minitab, *S-Plus* 和 *SYSTAT* 软件都可以作生存曲线，特别是 Kaplan-Meier 估计．在估计生存曲线时，有时我们需要求出删失数据的方差，在本书中没有介绍，但这里提到的软件包可以解决这些问题．

例 2.2.7

要测试 10 个汽车上风扇皮带的质量，我们把它们装到车上，并记录每辆车上皮带所能承受的里程数．测试结束后，5 个皮带都断裂了，寿命（以千英里计）分别为 77, 47, 81, 56, 80．另外 5 个没有断裂，分别是 62, 60, 43, 71, 37．生存函数的 Kaplan-Meier 估计如下所示．

i	u_i	结果	$\hat{p}_i$	$\hat{P}(u_i)$
1	37	缺失	10/10	1
2	43	缺失	9/9	1
3	47	死亡	7/8	0.875
4	56	死亡	6/7	0.75
5	60	缺失	6/6	0.75
6	62	缺失	5/5	0.75
7	71	缺失	4/4	0.75
8	77	死亡	2/3	0.5
9	80	死亡	1/2	0.25
10	81	死亡	0/1	0

对所有 $x>0$，$\hat{P}(x)$ 的图像如图 2-3 所示．

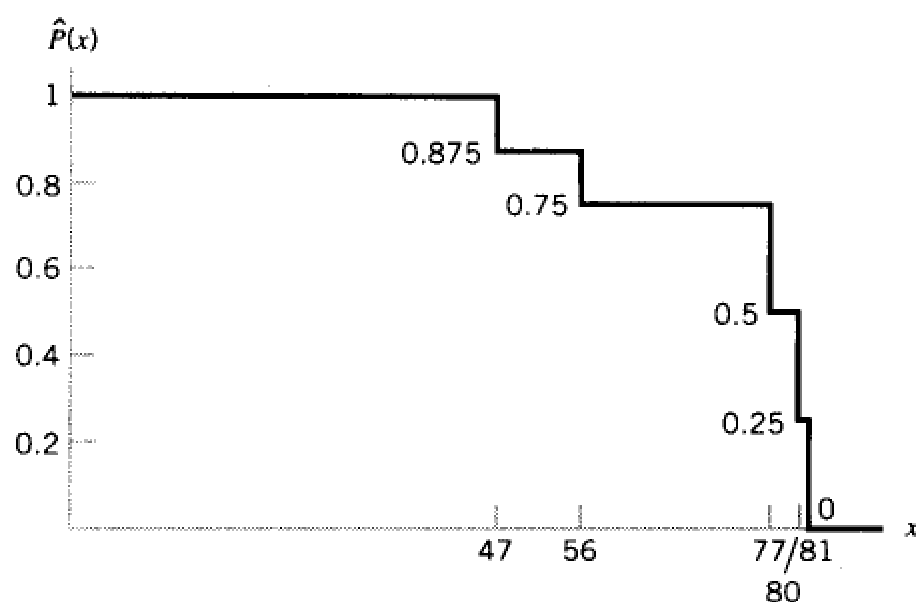


图 2-3 生存函数 $P(x)$ 的 Kaplan-Meier 估计 ■

如果没有缺失数据，只是死亡数据，那么 Kaplan-Meier 估计和 $1-S(x)$ 是一样的，从 $\hat{P}(x)=1$ 开始，在每个死亡时刻以 $1/n$ 为阶梯高度下降，直到 $\hat{P}(x)=0$. 如果既
91 有缺失数据又有死亡数据，$\hat{P}(x)$ 从 1.0 开始，下降的阶梯高度则不再一致. 如果在最后知道死亡时刻之后仍有缺失数据，那么 $\hat{P}(x)$ 不会下降到 0，并且对于那些在最后已知缺失数据后面的 x，它没有定义. 在这种情况下，通过本章前面介绍的常用方法，用 $S(x)$ 来估计 $F(x)$ 的某些相关参数，如均值，方差等是不太合适的，但是可以估计它的一些分位数. Kaplan 和 Meier(1958)也介绍了一些特殊的方法.

在这种意义下，点估计是一种非参数统计方法，因为不用了解任何关于未知分布函数的形式就可以做出点估计. 本章中的例子足以说明这一点.

很难说清构造置信区间的方法是参数的还是非参数的. 如果构造置信区间时不需要任何分布函数的形式，这显然是非参数方法，例 2.2.5 和 2.2.6 所示的近似方法就是非参数方法. 另一方面，如果方法要求未知分布函数是正态分布（见定义 1.5.3），或是其他的特殊形式，那么这种方法就是参数的. 我们将在 3.1,3.2,5.1,5.5,5.7 和 6.1 节中介绍其他几种构造置信区间的非参数方法.

习题

1. 从某一社区中随机抽取 10 名居民，他们中 5 个去年税前收入分别是(美元)8600,15 200,
92 16 200,16 400 和 29 600；而其他 5 个人没有收入.
   (a) 画出经验分布函数图像.
   (b) 求收入的样本中位数.
   (c) 求收入的样本均值.
   (d) 求收入的样本方差.
   (e) 求收入的样本标准差.
2. 在五场连续的比赛中，某篮球队分别得分 73,68,86,78 和 65.
   (a) 画出经验分布函数图像.
   (b) 求样本上四分位数.
   (c) 求样本四分位数极差.
   (d) 求样本均值.
   (e) 求样本标准差.

3. 随机抽取 5 个“12 盎司”的谷物盒子，其实际所盛谷物重量分别为 12.6,13.0,12.1,11.8 和 12.1 盎司.
   (a) 画出经验分布函数图像.
   (b) 求样本上下四分位数和四分位数极差.
4. 6 名学生的考试分数为 81,85,89,90,90 和 98.
   (a) 画出经验分布函数图像.　　(b) 求样本四分位数极差.
5. 用本章中相同的方法进行点估计，对于给定的 c，基于与 Y 有相同分布函数的随机样本 $X_1,X_2,\cdots,X_n$,求出概率 $P(Y\leqslant c)$的点估计. 换句话说，如果 $X_1,X_2,\cdots,X_n$ 是一组随机样本，其分布函数是 $F(x)$，估计 $F(c)$. 试估计习题 2 的下一场比赛中得分超过 80 分的概率是多少?
6. 用本节介绍的方法进行点估计，寻找估计量来估计随机变量的极差. 样本的极差会比总体极差更大吗? 会更小吗? 样本极差的期望值会比总体极差小吗?
7. 为检验一市立银行最小月度结余，从中随机抽取 175 家支票账户，样本均值为 1156 美元，标准差是 855 美元. 求该银行所有 14 000 家支票账户平均最小月度结余 90% 的近似置信区间.
8. 在一中心高中学生的学习能力测试研究中，随机抽取 50 名学生，平均得分为 81%，标准差为 11%. 求该中学 1159 名高中生平均分 95% 的置信区间.
9. 对健身俱乐部 18 名新成员进行体能测试. 用 X 表示他们中超重成员的比例，Y 记录蹬自行车锻炼后心律恢复情况. 二者样本相关系数为 $r=0.35$，为了估计所有新成员的过去、现在和将来的真实相关系数 95% 的置信区间，我们使用自助法从小到大排列 300 个自助样本，得到最小的 10 个 r^* 值分别是 93

   $$-0.15,\ -0.06,\ -0.02,\ 0.01,\ 0.03,\ 0.03,\ 0.05,\ 0.06,\ 0.07,\ 0.09$$

   最大的 10 个 r^* 值分别是

   $$0.51,\ 0.53,\ 0.53,\ 0.55,\ 0.56,\ 0.57,\ 0.59,\ 0.59,\ 0.60,\ 0.62.$$

   求总体相关系数 95% 的近似置信区间.
10. 设上面第 9 题中 300 个自助样本 r^* 的样本均值为 0.30，且标准差为 0.12，那么 r 的标准误差的估计值是多少? 当样本容量为多少时该估计有效?
11. 为了确定一批灯泡的生存函数，从中随机抽取 8 个灯泡做测试. 测试是这样进行的：先用 4 个灯泡做试验，若其中任意一个熄灭，则用剩下 4 个中的任一个灯泡代替. 这样，烧坏的灯泡的寿命分别是 187,196,206,210,273 小时. 试验终止时没烧坏的 3 个灯泡分别亮了 127,190,194 个小时. 用 Kaplan-Meier 估计量来估计该批灯泡的生存函数.
12. 1997 年 9 月，100 名学生参加了一项关于牙齿卫生的 2 年计划项目. 72 名学生完成了第 1 年的计划，55 名学生完成了这个 2 年计划项目. 1998 年 9 月，又有 100 名学生加入了该项目，57 名学生完成了第 1 年的计划. 用 Kaplan-Meier 估计量来估计 1999 年 9 月加入该项目的学生能够完成这个 2 年计划项目的概率.

思考题

1. 因为估计量是随机变量，那么如果给出足够的信息，我们就能得到它的概率分布. 假设一个有限总体由 4 个元素构成，测量值分别为 4,6,7 和 10. 从该总体中无放回地抽取容

量为2的随机样本.

(a) 随机样本共有多少种可能?

(b) 列举所有可能的样本.

(c) b中每组样本出现的概率是多少?

(d) b中每组样本的样本中位数是多少?

(e) 在d中抽到样本中位数的概率是多少?

(f) 画出样本中位数的分布函数图像.

(g) 用上面列出的方法求出样本极差的概率函数.

2. 一个统计量称为总体参数的无偏估计量，如果满足它的期望等于被估参数.

(a) 求思考题1中的样本中位数的期望. 它和总体中位数相等吗? 样本中位数是总体中
94 位数的无偏估计吗?

(b) 求思考题1中的样本极差的期望. 它和总体的极差相等吗? 样本中位数是总体极差的无偏估计吗? (和习题6做对比.)

2.3 假设检验

统计推断有很多形式，其中在非参数方法中，研究者和应用者广为接受和关注的是假设检验，在本节和下一节中会详细介绍.

假设检验是根据样本来推断总体的一些给定陈述是否成立的过程. 这些陈述称为假设，下面是几个包含陈述假设的例子：

1. 女人比男人更易发生机动车交通事故.
2. 上托儿所能够帮助孩子在小学学习中取得更好的成绩.
3. 被告有罪.
4. A牙膏在防蛀方面比B牙膏更有效.

特殊假设的非统计检验是很容易进行的. 我们可以观测一批和假设相关的数据，或是不相关的一批数据，或是根本没有数据，然后得出接受或拒绝假设的结论，尽管这个结论是可疑的. 但我们所要讨论的假设检验的类型是比较合理的，它称为统计假设检验，检验的过程有着合理的定义. 这里给出了这种检验的几个简要步骤：

1. 假设是根据总体提出的，其中包含两种假设. 试验者希望证实的假设称为备择假设 (alternative hypothesis)，或者在质量控制中，它是指关于产品或服务质量不令人满意或"失控"的一些陈述. 典型的备择假设为"新产品比旧产品要好，"或"这种药对治这种病更有效."，有时，备择假设也指研究假设 (research hypothesis).

和备择假设对立的称为零假设 (null hypothesis) 或检验假设 (test hypothesis). 这是在假设检验中需要被检验的假设. 上面的例子中和备择假设相对应的零假设分别是"新产品不比旧产品好"，"这种药对治疗这种病不是更有效". 在质量控制中，零假设的陈述是指关于产品或服务质量让人满意的一些陈述.

95 如果样本数据强有力地与零假设不一致，那么拒绝零假设. 如果样本数据和零假

设不矛盾，或是没有充足的理由显示数据和零假设有冲突，那么试验者“不能拒绝”零假设. 有时试验者也说“接受零假设”，它所表达的和前面是同一个意思，该叙述不能误解为数据证明零假设是真的. “接受零假设”只是表示不能拒绝零假设.

2. 选择检验统计量（test statistic）. 一个好的检验统计量在零假设成立时取一些值，而在零假设不成立时取另外一些值. 也就是说，一个好的检验统计量在判断数据是否和零假设一致方面是个敏感的指标.

3. 根据检验统计量的可能取值，构造是否接受零假设的决策法则（decision rule）.

4. 基于从总体中抽取的随机样本，从而得到检验统计量的取值，最后做出是否接受零假设的判决.

下面的例 2.3.1 更加精确地描述了上面假设检验的这个过程.

例 2.3.1

某机器生产零件，当次品率等于或低于 5% 时可以认为该机器工作正常；高于 5% 时，就需要对机器引起注意. 零假设为

$$H_0\text{: 该机器正常工作}$$

是一个要检验的假设. 备择假设为

$$H_1\text{: 需要注意该机器}$$

如果 H_1 是真的，它就是我们要能检测的假设. 从该机器生产的所有零件中随机抽取 10 个，根据这组随机样本检验 H_0. 如果拒绝 H_0，我们需要采取修理措施来使机器正常工作.

假设每个零件是次品的概率均为 p，且是否为次品相互独立. 因此，在这个假设模型中，原来的假设 H_0 与 H_1 等价于

$$H_0: p \leqslant 0.05$$
$$H_1: p > 0.05$$

我们知道，如果次品太多，就要拒绝 H_0. 所以令检验统计量 T 为次品的总个数，那么，根据例 1.3.5，T 服从参数为 p，n 为 10 的二项分布. 由表 A3 我们看到，若 H_0 96
为真（$p \leqslant 0.05$），那么

$$P(T \leqslant 2) \geqslant 0.9885 \tag{1}$$

当 $p = 0.05$ 时取等号，且

$$P(T > 2) \leqslant 0.0115 \tag{2}$$

当 $p = 0.05$ 时取等号. 由于当 H_0 为真时拒绝 H_0 的概率很小，即小于等于 0.0115，所以我们决定，若 T 超过 2，则拒绝 H_0. 样本空间中对应于 T 大于 2 那些样本点的集合称为临界域（critical region）. 决策法则为：若观测结果在临界域中（T 超过 2），则拒绝 H_0；否则接受 H_0.

假设 10 个零件的随机样本中有 4 个次品，那么 $T = 4$，拒绝零假设，则我们认为需要注意该机器的工作状况. ■

在例 1.3.1 中，针对收集数据和这种数据的类型条件作出了一些假设. 在构造模型和理想化试验时，这些假设是等价的. “在这个模型下” 意味着 “在这些假设下”，试验者则尽可能地在满足这些假设的条件下收集数据.

在这个模型下，原来的假设用统计术语可以重新叙述为另一种等价的形式. 这些假设可分为简单（simple）假设和复合（composite）假设.

定义 2.3.1 若假设为真，则在样本空间中只定义了一个概率函数，这时称该假设为简单假设. 若假设为真，则在样本空间中定义了两个或更多个概率函数，这时称该假设为复合假设.

在这个例题中，模型在每个样本点导出了二项概率 $\binom{10}{k}p^k(1-p)^{10-k}$，其中样本点对应 k 个次品和 $10-k$ 个非次品. 这表示 p 决定了定义在样本空间中的一组概率函数.（对于每个样本点，k 已知）假设 H_0 为真，但 p 可能是 0 到 0.05 中任一值，则有几种可能的概率函数，且 H_0 为复合假设. 对于 H_1，有同样的结论. 假设 “$p=0.05$” 是一个简单假设，事实上，若 $p=0.05$ 为真，那么概率函数将表示 k 个次品的样本点赋予概率 $\binom{10}{k}(0.05)^k(0.95)^{10-k}$，这里合理地定义了概率函数（不含未知参数），且只有一种可能.

97 **定义 2.3.2** 一个统计检验量是指在假设检验中能够帮助作出判决的统计量.

临界域

一个好的检验统计量应该具备这样的理想性质：它把样本空间中的点和实数对应起来，该样本空间中的样本点是按照区分零假设 H_0 是否为真的能力来排列的. 例如，检验统计量给那些最能够帮助试验者决定拒绝 H_0 的样本点赋予较大的值，给那些帮助试验者决定接受 H_0 的样本点赋较小的值，那么检验统计量的值越大，试验结果表明越应该拒绝 H_0，这样当检验统计量的所有值比某一个数都大时，则应拒绝 H_0. 进一步讲，这能够使试验者不论拒绝域多么大还是多么小，都能客观地得出相同的结论. 拒绝域对应检验统计量中最大值的检验称为*右边检验*（upper-tailed test）. 同样，若次序相反，那么拒绝域对应检验统计量中最小值的检验称为*左边检验*（low-tailed test）.

这两个都是*单边检验*（one-tailed test）. 例题中的检验就是单边的. 若拒绝域对应检验统计量中最大值和最小值，那么该检验称为*双边检验*（two-tailed test），因为拒绝域对应于检验统计量可能的两个 “边”.

定义 2.3.3 *临界域*（critical region）是样本空间中使得拒绝零假设全体样本点的集合.

有时临界域亦称为*拒绝域*（rejection region），所以很明显样本空间中不在临界域的全体样本点的集合称为*接受域*（acceptance region）.

错误类型

在假设检验中有可能作出两种类型的错误判决. 如果零假设为真, 而我们错误地拒绝了它, 那么我们所犯的错误是第一种错误 (error of the first kind), 亦称第一类错误 (type Ⅰ error). 也就是说, 当 H_0 为真, 而我们试验的结果却落在临界域内时, 即发生了第一类错误.

定义 2.3.4 第一类错误是拒绝了正确零假设的错误.

假设检验中另外一类错误是指当零假设为假时, 却接受了零假设, 这类错误是第二种错误 (error of the second kind), 亦称第二类错误 (type Ⅱ error). 98

定义 2.3.5 第二类错误是接受了不正确零假设的错误.

显著性水平

这两类错误可以和一定的犯错误概率联系在一起, 首先考虑犯第一类错误的概率.

定义 2.3.6 显著性水平 (level of significance) α 是拒绝正确零假设的最大概率.

显著性水平可以这样求得: 首先假设零假设 H_0 成立, 然后确定一样本点落入临界域的概率. 如果 H_0 是简单假设, 那么 H_0 成立只产生一个定义在样本空间上的概率函数, 则 α 是把临界域中所有点的概率加到一起的总和. 但是通常在假设 H_0 为真时, 通过计算检验统计量取某个值的概率来确定 α 会更容易些, 而这个值应导致拒绝 H_0.

零分布

在统计假设检验中, 了解在零假设成立时检验统计量的概率分布是非常必要的, 这称为检验统计量的零分布 (null distribution).

定义 2.3.7 检验统计量的零分布是当零假设成立时, 检验统计量的概率分布.

在例2.3.1中, 检验统计量 T (即10个零件中次品的个数) 的零分布是参数 $p\leqslant 0.05$ 的二项分布, 这是由于我们假设了独立性和概率 p 是常数. 每个统计假设检验的显著性水平都可以由检验统计量的零分布得到.

如果 H_0 是一复合假设, α 是拒绝 H_0 的最大 (maximum) 概率, 这里的最大值是当零假设成立时, 所考虑的概率分布可能值的最大值. 在这个例题中, H_0 是复合的, 那么对每个不同的 p 值, 拒绝正确零假设的概率为

$$\begin{aligned} P(\text{拒绝} H_0) &= P(T>2 \mid H_0 \text{为真}) \\ &= \sum_{i=3}^{10}\binom{10}{i}p^i(1-p)^{10-i}; \qquad p\leqslant 0.05 \end{aligned} \tag{3}$$

99 但式3中的概率当 p 取最大值时，它达到最大值. 在 H_0 下，p 的最大值是0.05，所以由表A3或式2，显著性水平由下式给出

$$\begin{aligned}\alpha &= \max P(T>2 \mid H_0 \text{为真}) \\ &= P(T>2 \mid p=0.05) \\ &= 0.0115 \end{aligned} \tag{4}$$

很显然，显著性水平有时称为*临界域的大小*（size of the critical region）. 因为，若 H_0 成立，拒绝 H_0 的最大概率是 α，则接受 H_0（即作出正确判决）的最小概率是 $1-\alpha$.

犯第二类错误的概率用 β 表示. 显然在假设检验中我们希望 α 和 β 都接近于零. 在实际应用中，样本容量可以帮助我们决定 α 和 β 会有多小. 只有当样本包含了总体所有的信息时，犯错误的可能性才可能被完全消除.

功效

假设 H_0 为假，接受 H_0 的概率是 β，或是拒绝 H_0 的概率是 $1-\beta$，后面这一概率表示了该检验检测错误零假设的*检验功效*（power of the test）.

定义 2.3.8 *功效*（power）是拒绝错误零假设的概率，记为 $1-\beta$.

与 α 不同，功效不总是唯一的. 如果 H_1 是简单假设，那么由 H_1 成立（等价于"H_0"为假）所导出的概率函数只有一个，即一个拒绝 H_0 的概率，或得到一个落入临界域的样本点. 因此这时 $1-\beta$ 唯一. 如果 H_1 是复合假设，那么在 H_1 下的每一个概率函数都会有不同的 $1-\beta$ 值，这时，功效取决于多个不同可能的概率函数.

		决定	
		接受 H_0	拒绝 H_0
真实情形	H_0 正确	正确判决 概率 = $1-\alpha$	第一类错误 概率 = α(显著性水平)
	H_0 错误	第二类错误 概率 = β	正确判决 概率 = $1-\beta$(功效)

100

前面已经讨论了错误的类型，现在我们转向讨论临界域. 尽管我们已经讨论了一些有关临界域的内容，但并没有涉及到它是如何选取的. 如果检验统计量已经选定，并由它确定了单边或是双边检验，那么临界域的选择只取决于试验者对临界域大小，即显著性水平的偏向. 通常，显著性水平 α 的减小会伴随着 β 的增加，在假设检验中，我们的两个目标是：若 H_0 为真，那么以最小的可能性拒绝 H_0；若 H_0 为假，那么以最大的可能性拒绝 H_0. 所以在那些有固定大小的 α 的点集中，临界域通常是那些 $1-\beta$ 最大值所对应的样本点的集合. 习惯上，α 通常取0.05或0.01，并且临界域还要根据检验统计量的可能值来确定.

检验的 p-值

如果引入检验的 p-值（p-value），假设检验的结果会更有意义.

定义 2.3.9 检验的 p-值是根据已知观测，零假设被拒绝时的最小显著性水平.

令 t_{obs} 表示检验统计量 T 的观测值. 在右边单边检验中，p-值是由 T 的零分布计算得到的 $P(T \geqslant t_{obs})$ 值. 在左边单边检验中 p-值是 $P(T \leqslant t_{obs})$.

在双边检验中，p-值规定为单边检验中两个 p-值中较小值的 2 倍. 严格来讲，如果 T 的零分布是离散的，并且拒绝域的右边和左边概率不相等，这不太可能在两边构造概率相等而精确的显著性水平. 所以这和前面的定义是不一致的. 但是，为了避免定义模糊，我们在后面还是认为双边的 p-值是观测值落在零分布单边概率的 2 倍.

例 2.3.1 中的检验是右边的，T 的观测值是 4，所以由表 A3 可知 p-值为 $P(T \geqslant 4 \mid p=0.05)=0.0010$. p-值有时简写为 p，但是在例 2.3.1 中这个符号表示次品的概率，所以这里最好用"p-值"以免混淆.

在许多发表的研究结果中，统计检验浓缩为只包括检验的名称，假设和 p-值的报告. 若 p-值小于或等于 α，则拒绝零假设，这里 α 通常取 0.05.

例 2.3.2

为了检验上过和没上过幼儿园的孩子在学习上是否表现不同，选择 12 个三年级的 101
学生进行研究，其中 4 个上过幼儿园. 要检验的零假设是

H_0: 三年级学生学习上的表现不取决于他们是否上过幼儿园

备择假设是

H_1: 学习上的表现和上过幼儿园之间是不独立的

模型假设这 12 个孩子是所有三年级学生中的一组随机样本，并且根据学习成绩（从好到差）把这些孩子从 1 到 12 排序标记. "不独立"是指上过幼儿园的孩子整体比没上过幼儿园的孩子表现好，或整体表现不好. 在这个模型下，假设可以重新叙述为

H_0:上过幼儿园的 4 个孩子的秩是秩 1 到 12 的一个随机样本

H_1:上过幼儿园的 4 个孩子的秩整体比 12 个孩子中随机抽取 4 个孩子的秩要大或小

我们选择一检验统计量 T，是上过幼儿园的 4 个孩子的秩和. 我们令那些与很大或很小的 T 值对应的样本点构成拒绝域，所以该检验是双边的.

每一个可能的结果是从 1 到 12 中抽取的 4 个数，且对应着上过幼儿园的 4 个孩子的秩，所以样本空间中有 $\binom{12}{4}=495$ 个点. 为了决定临界域包含哪些点，我们将假设 H_0 为真，并且在决定临界域时，看一下 α.

如果 H_0 为真，4 个孩子的秩应当是 12 种可能中的一组随机样本，因此每 4 个秩的选择都是等可能性的，这样样本空间中的每个点概率相等，它为 1/495. 这样 H_0 是一简单假设. 因为我们决定用双边检验，所以看一下 T 较大和较小值所对应的样本点，T 可能的最大值和最小值是 42 和 10，对应的样本点分别是（12,11,10,9）和（1,2,3,4）. T 其他的大值和小值所对应的试验结果如下所示：

102

T	样本点	T	样本点
10	(1, 2, 3, 4)	42	(9, 10, 11, 12)
11	(1, 2, 3, 5)	41	(8, 10, 11, 12)
12	(1, 2, 3, 6)	40	(7, 10, 11, 12)
12	(1, 2, 4, 5)	40	(8, 9, 11, 12)
13	(1, 2, 3, 7)	39	(6, 10, 11, 12)
13	(1, 2, 4, 6)	39	(7, 9, 11, 12)
13	(1, 3, 4, 5)	39	(8, 9, 10, 12)
14	(1, 2, 3, 8)	38	(5, 10, 11, 12)
14	(1, 2, 4, 7)	38	(6, 9, 11, 12)
14	(1, 2, 5, 6)	38	(7, 8, 11, 12)
14	(1, 3, 4, 6)	38	(7, 9, 10, 12)
14	(2, 3, 4, 5)	38	(8, 9, 10, 11)

注意到有 12 个样本点对应于 $T\leqslant 14$，有 12 个样本点对应于 $T\geqslant 38$. 假如临界域由所有 $T\leqslant 14$ 或 $T\geqslant 38$ 对应的样本点组成，则 α 为

$$\alpha=\frac{\text{临界域中的点数}}{\text{样本空间中的点数}}$$

$$=\frac{24}{495}=0.0485 \tag{5}$$

因为在 H_0 下，样本空间中所有的样本点的概率相等. 我们的决策法则是：若 T 的观测值$\leqslant 14$ 或$\geqslant 38$，我们拒绝 H_0；否则我们接受 H_0.

经过观测，上过幼儿园的孩子在 12 个孩子中学习成绩的排序分别是 2,5,6 和 9，得到 T 值为：

$$T=22 \tag{6}$$

所以我们接受 H_0. 由正态分布可以得到 p-值的近似值（参考例 1.5.7）. 左边 p-值是当零假设成立时，$T=22$ 或更小值的概率. 由定理 1.4.5 可以得到 T 的均值和方差，分别为 26 和 34.67（$n=4, N=12$），所以 T 的标准差为 5.888. 由表 A1，正态近似为

$$P(T\leqslant 22)\cong P\left(Z\leqslant\frac{22-26}{5.888}\right)=P(Z\leqslant -0.6794)=0.248 \tag{7}$$

它的 2 倍则是双边检验 p-值 0.496.

这么大的 p-值表明了在零假设成立时，T 的观测值是所期望的，因此由数据，我们没有理由怀疑零假设不正确. ■

例 2.3.2 中所示的检验过程叫做 Mann-Whitney 检验或 Wilcoxon 检验. 我们将在

第5章就它的多种形式进一步讨论. 例2.3.2中的数据采用的是度量的次序尺度，我们不需要知道每个孩子学习成绩的具体数值，事实上，这种具体数值所反映的信息通常没有什么价值，因为每个学校，甚至每个老师对这些数值都有不同的解释和标准，而这种排序则有通用的解释.

例2.3.1给出了对名义数据的分析，“次品”或“非次品”. 例2.3.1中的检验是基于二项分布的，这种检验和其他基于二项分布类型的检验将在第3章中正式介绍. 103

计算机辅助

绝大多数的统计软件包都能够做假设检验. 在一些软件包中，使用者指定零假设和备择假设，然后该软件包给出 p-值. 而在其他的软件包中，计算机总是给出双边检验的 p-值，使用者必须决定该值是否为我们要求的，或是必须取其一半的值而得到一单边 p-值. 若 p-值小于等于使用者给定的显著性水平，那么则拒绝零假设.

很多计算机软件包使用近似方法来求 p-值. 大多数情况下这是可行的，但是并不是所有情形都可行. 越来越多的计算机软件包在仿照 *StatXact* 的例子，它计算精确的 p-值，或当精确的 p-值在实际中不能得到时，运用蒙特卡洛模拟法得到近似的 p-值.

习题

1. 检验一种新的教学方法是否比现行的教学方法更好.
   (a) 合适的 H_0 和 H_1 分别是什么?
   (b) 问题中“显著性水平”表示的是什么?
   (c) 问题中“功效”表示的是什么?
2. 法官审判被告，在证明被告有罪前，假设被告是无罪的.
   (a) 谁在做假设检验?
   (b) H_0 和 H_1 分别是什么?
   (c) 样本和总体是什么?
   (d) 问题中的“显著性水平”和“功效”分别意味着什么?
3. 对于下面的每一项，合适的 H_1 是什么?
   (a) H_0: 肥料 B 至少和肥料 A 一样好.
   (b) H_0: 我的对手没有作弊.
   (c) H_0: 太阳黑子的出现不会影响经济周期.
4. 对于下面的每一项，合适的 H_0 是什么?
   (a) H_1: 该研究对象有超感知觉.
   (b) H_1: 探测杆在发现水源中很有作用.
   (c) H_1: 年平均气温正在上升. 104
5. 一枚硬币掷5次，记录正面或反面出现的观测结果. 临界域是“至少4次出现正面”的事件. 假如 H_0 正确，那么样本空间的所有的样本点概率相等. α 是什么? 假如 H_1 成立，每次投掷时，“正面”出现的概率为0.6，功效是多少?

6. 一枚硬币掷 4 次，临界域是“不多于 1 次正面”的事件. 令 $p=P$（正面出现）. 假设 H_0: $p=0.5$ 和 $H_1: p=0.1$. 求该检验的功效. 这个问题中还应该做出哪些题目中未提及的其他的假设?
7. 样本空间包含 10 个样本点，临界域内只有一个样本点. 假如 H_0 成立，那么所有的样本点概率相等. 假如 H_1 成立，临界域中的样本点概率为 0.91，其他每个样本点的概率是 0.01，α 是什么? 功效是多少?
8. 假设样本空间包含 50 个样本点，其中 2 个样本点分别命名为 A 和 B. 假如零假设成立，则所有的样本点概率相等. 假如备择假设成立，则样本点 A 和 B 发生的概率是其他 48 个样本点的 26 倍，而这 48 个样本点等概率. 临界域由 A 和 B 组成.
   (a) 求出显著性水平 α.　　(b) 求出功效.

思考题

1. 罐子中有 12 个塑料筹码，从 1 到 12 连续标号. 该试验是从罐中有放回地随机抽取 2 个筹码，试验结果由 2 个筹码上的号码按照抽取次序构成. 令检验统计量 X 为抽取的筹码上的数字总和，临界域是由 X 值小于 5 的样本点构成的. 假如 H_0 成立，则筹码的抽取是随机的，如果假如 H_1 成立，则抽取筹码 1,2,3 的概率是抽取其他筹码概率的 2 倍.
   (a) 列举临界域中的样本点.　　(b) 求出 α.
   (c) 求出功效.　　(d) H_0 和 H_1 是简单假设还是复合假设?
   (e) 是单边检验还是双边检验?
2. 7 个筹码从 1 到 7 连续标号，独立地放到 A 和 B 两个盒子里. 试验结果由盒子 A 中的筹码号码组成，不计放入的次序. 令检验统计量 X 是盒子 A 中的筹码标号的总和，临界域是由 X 值小于 6 的样本点构成的. 假如 H_0 成立，则每个筹码放入盒子 A 中的概率为 0.5，假如 H_1 成立，则该概率为 0.3.
   (a) 列举临界域中的样本点.　　(b) 求出 α.
   (c) 求出功效.　　(d) H_0 和 H_1 是简单假设还是复合假设?
105 (e) 是单边检验还是双边检验?

2.4　假设检验的性质

假设一旦确定后，对于检验零假设通常有几种假设检验方法. 为了从中选择一种方法，我们要仔细考虑这些检验的一些性质，其中最重要的一个问题是：“这个检验的假设条件适用于我的试验吗?”如果答案是“不适用”，那么我们可能不能用这个检验. 但是，在舍弃这个检验前，应该明确检验背后的假设条件. 例如，大多数参数检验中所做的一个假设是被检验的随机变量服从正态分布，进一步研究表明，若随机变量的分布只要稍微与正态分布有相似之处，检验仍近似有效. 所以隐含的假设是“近似正态”，并且若假设条件是“近似成立”的，那么该假设不应该舍弃. 但是，该检验的不足之处应有所记录. 另外一个准则就是在模型中相对于有较多假设条件的检验，我们更喜欢有较少假设条件的检验.

有两个原因说明检验的假设条件不满足时，我们仍然使用该检验是危险的. 首先，拒绝零假设不是因为由数据指出的零假设是错误的，而是由于数据表明检验的其中一个假设条件不成立. 第二个危险是，有时数据明显地表示零假设是错误的，并且模型中一个错误假设也影响着数据，但是在检验中，这两种影响相互抵消了，所以这个检验什么也没揭示就接受了零假设. 一般的假设检验不仅对错误的假设敏感，同时对模型中错误的假设条件也一样灵敏.

基于前面的准则，我们从适合的检验中，根据检验的其他性质来选择最好的检验. 本节将在后面对有关性质具体定义，它们是

1. 检验应是无偏的.
2. 检验应是相合的.
3. 在某种意义上，检验应比其他的检验更有效.

其中，最重要的也是被广泛应用的是有关功效的有效性.

有时，一个检验能满足上面三条标准中的一两条，我们就很满意了. 很少有三条能够同时满足的. 本节后面将要讨论检验的无偏性、相合性、有效性和检验的功效.

功效函数

若 H_1 是复合假设，功效随着概率函数的变化而变化. 如果 H_1 是按照某些未知参数来陈述的，那么功效通常作为该参数的函数形式给出，这种函数称为*功效函数* 106 (power function)，可用代数形式和图像来表达. 功效是当 H_1 成立时拒绝 H_0 的概率，和功效不一样，功效函数通常是对在 H_0 和 H_1 下参数的所有值而定义的. 这样说来，功效函数比功效给了我们更多的信息，它是当 H_0 成立或不成立时，拒绝 H_0 的概率.

例 2.4.1

例 2.3.1 中的临界域是由 10 个抽样产品中多于 2 个次品的所有样本点组成的. 在模型的假设下，样本点落到临界域的概率，同拒绝 H_0 的概率相等，即

$$P(\text{拒绝}\, H_0) = \sum_{i=3}^{10} \binom{10}{i} p^i (1-p)^{10-i} = 1 - \sum_{i=0}^{2} \binom{10}{i} p^i (1-p)^{10-i} \tag{1}$$

这里 p 是次品率. 拒绝 H_0 的概率是 p 的函数，由表 A3 可以画出该功效函数的大致图像.

p	$P(\text{拒绝}\,H_0)$	p	$P(\text{拒绝}\,H_0)$
0	0.0000	0.50	0.9453
0.05	0.0115	0.55	0.9726
0.10	0.0702	0.60	0.9877
0.15	0.1798	0.65	0.9952
0.20	0.3222	0.70	0.9984
0.25	0.4744	0.75	0.9996
0.30	0.6172	0.80	0.9999
0.35	0.7384	0.85	1.0000
0.40	0.8327	0.90	1.0000
0.45	0.9004	1.00	1.0000

如图2-4所示，零假设陈述了p在0到0.05之间. H_0成立时，图2-4中曲线的最大值是显著性水平，由前面式2.3.4计算等于0.0115. 功效的取值范围是从0.0115(p约为0.05)到1.0000(p等于1.0). ■

根据它们的功效函数可以比较这两种检验，这种比较的基础在本节后面定义了相对效率后再作讨论.

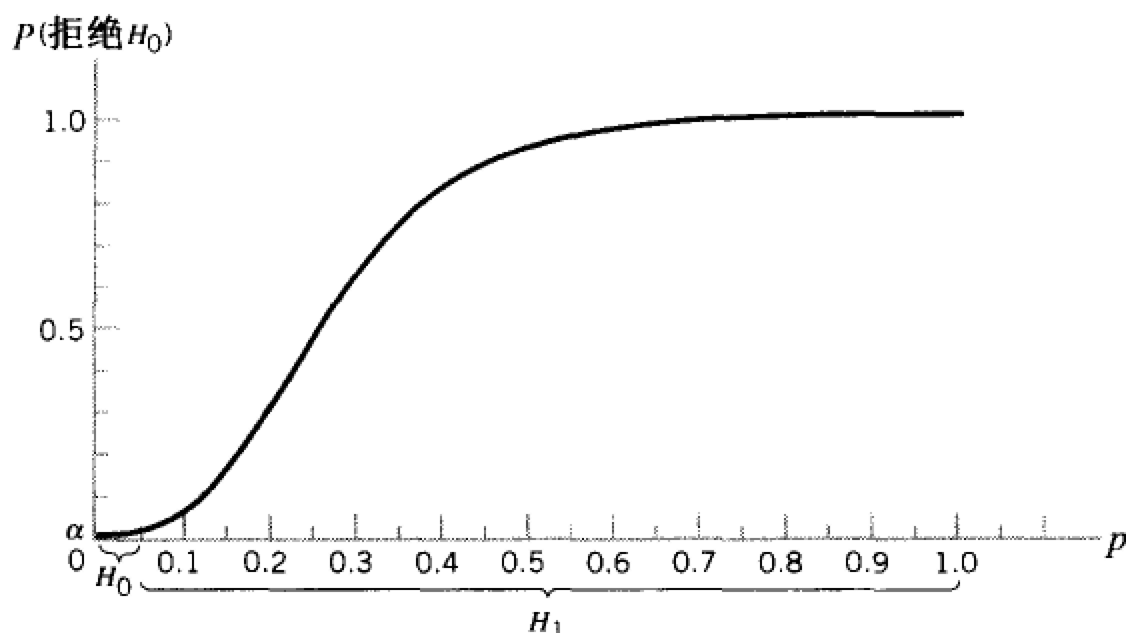


图2-4 功效函数

计算机辅助

检验的功效是显著性水平，所论的简单备择假设和样本容量的函数. 计算机软
107 件包*PASS*在给定显著性水平，备择假设中参数的取值范围和样本容量后，可以计算检验的功效. 在给定功效时，它也可以计算出要求的样本容量. *Minitab*也可以计算一些非参数检验的功效.

无偏检验

显然，我们希望拒绝H_0的可能性，在H_0不成立时要比H_0成立时大.

定义2.4.1 无偏检验(unbiased test)是H_0不成立时拒绝H_0的概率大于等于H_0成立时拒绝H_0的概率的检验.

因此无偏检验的功效至少和显著性水平一样大. 一个检验不是无偏的则称为有偏检验(biased test). 例2.3.1中描述的检验和本节例2.4.1中进一步讨论的检验都是无偏检验，这从图2-4中显然可以看出.

相合检验

检验的另一个优良性质是相合性(consistent). 虽然我们说一个检验是"相合的"或是"不相合的"，其实这里的相合是针对一系列检验而言的，因为它是当样本容量趋于总体容量时所使用的. 为方便起见，无论总体容量有限还是无限，我们都

将称总体容量“无限”. 从技术上讲，因为样本空间和临界域是随着样本容量的改变而改变的，所以对于每个不同的样本容量，我们都得到一个不同的检验. 因此，随着容量的增加，我们考虑一个检验序列，每一个样本容量都对应一个检验. 108

定义 2.4.2 称一检验序列对 H_1 中所有备择假设是相合的，如果对于 H_1 下的每一个可能固定的备择假设，当样本容量趋于无穷时，检验的功效趋于 1.0. 而序列中每个检验的显著性水平，尽可能地趋于但不超过某一固定的显著性水平值 $\alpha>0$.

例 2.4.2

我们要检验是否某一性别的婴儿出生率较高，而不是男婴和女婴出生概率相等，即要检验

H_0: 男婴和女婴出生概率相等

备择假设是

H_1: 男婴比女婴出生的概率大，或者小

抽样总体由某一国家登记的新生婴儿构成. 对于给定的 n 值，样本由最后登记的 n 个婴儿构成. 我们假设这种抽样方法在尽可能考虑到性别特征的情况下等价于随机抽样，假定男婴出生的概率 p 是常数，并且事件生“男”和生“女”是相互独立的，那么这些假设等价于如下

$$H_0: p = 1/2$$
$$H_1: p \neq 1/2$$

令检验统计量 T 为出生男婴的数目. 临界域选择为对称地对应到 T 的最大和最小值，分别称为 T 的右边和左边，最大显著性水平不超过 0.05.

因此，我们就给出了整个检验序列，其中每个检验对应着样本容量的每一个取值，都是双边的，显著性水平为 0.05 或更小，且 T 服从二项分布. 对于各种检验，Dixon(1953)给出的临界域如下：

n	T值对应的临界域			α
5		无		0
6	$T=0$	和	$T=6$	0.03125
8	$T=0$	和	$T=8$	0.00781
10	$T\leqslant 1$	和	$T\geqslant 9$	0.02148
15	$T\leqslant 3$	和	$T\geqslant 12$	0.03516
20	$T\leqslant 5$	和	$T\geqslant 15$	0.04139
30	$T\leqslant 9$	和	$T\geqslant 21$	0.04277
60	$T\leqslant 21$	和	$T\geqslant 39$	0.02734
100	$T\leqslant 39$	和	$T\geqslant 61$	0.03520

109

注意，对于所有 $n\leqslant 20$ 和由表 A3 得到的值是相同的. 对于 $n>20$，可以用正态近似（例 1.5.6）的结果，但是表中精确的结果可能更好.

为了看检验序列是否相合，我们来比较这些检验的功效函数. 由Dixon(1950)给出的表，我们在图2-5中画出了其中的几个功效函数. 我们可以看到随着样本容

量的增加，对于每个固定的p-值（除了$p=0.5$），功效一直增大到1.0.

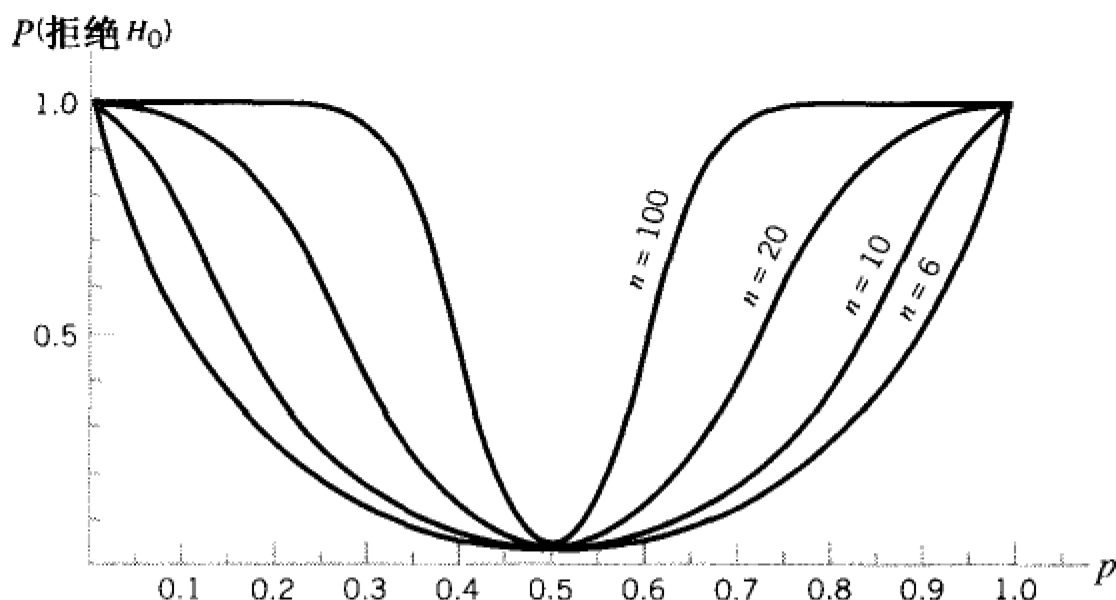


图2-5 几个功效函数的比较 ■

这个例子只是表明检验序列相合性背后的思想，而不是严格地证明检验序列是相合的. 相合性严格的证明通常需要更多的数学理论，而在本书中，我们不涉及这部分内容，所以只是给出一检验序列（或一"检验"）是否相合的结论.

相对效率

我们已经定义了统计检验的许多其他性质，相关内容在各种著作中（如，Lehmann,1959）都可以找到. 我们再讨论一个性质，即*效率*（efficiency）. 效率是一个相对的术语，它是用来比较在相同条件下两种检验的样本容量. 假定有两个检验用来检验特定的H_0对H_1，而且它们有相同的α和β值，因此关于显著性水平和功效它们是"可比的".（注意，两种检验的β值是相等的，通常排除了备择假设是复合假设的情况，因为这时通常β不只有一个值.）需要的样本量越小的检验越好，因为小
110 样本量意味着试验中用更小的花费和更少的精力. 较小的样本量的检验被称为比其他检验*更有效*（more efficient），*相对效率*（relative efficient）也更大.

定义 2.4.3 设T_1和T_2分别表示两种检验，用来检验相同的H_0对H_1，临界域对应的α和β相等，T_1对T_2的*相对效率*（或"T_1相对于T_2的效率"）定义为比值n_2/n_1，其中n_1和n_2分别是检验T_1和T_2的样本容量.

根据定义2.4.3，如果n_1小于n_2，T_1相对于T_2的效率比1大，和我们预想的一样.

假如备择假设是复合的，相对效率可以由备择假设定义的每个概率函数计算得到，这些相对效率值可用表格，或有时用图像来表示.

例 2.4.3

在相同的H_0对H_1下，两种检验有相等的$\alpha=0.01$和$\beta=0.14$. 第一个检验的样本量为75，第二个为50. 因此第一个检验不如第二个检验有效. 第一个检验对第二个检验的相对效率为

$$\frac{50}{75} = 0.67$$

第二个检验相对第一个检验的效率为

$$\frac{75}{50} = 1.5$$

若已知 $\alpha = 0.05, \beta = 0.30, n_1 = 40$，第一个检验相对于第二个检验的效率是 0.75，那么可以得到要求的第二个检验的样本容量.

$$相对效率 = \frac{n_2}{n_1}$$

$$0.75 = \frac{n_2}{40}$$

$$n_2 = 30$$

第二个检验方法用 30 个样本，就能够达到第一个检验用 40 个样本得到的一样好的分析结果. ■

渐近相对效率（A. R. E）

相对效率依赖于 α 的选择，β 的选择，以及复合假设 H_1 中的特定备择假设，为了提供一个检验与其他检验进行全面的比较，相对效率显然依赖于太多的参数. 我们更希望进行比较而不依赖于 α, β 以及当 H_1 是复合假设时，H_1 中特定备择假设的选择. 有时这种方法可以简要叙述如下. 111

考虑一检验序列，对于同一固定的 α，假如检验序列相合，那么随着样本量 n_1 的增加，β 变小. 为了不让 β 变小，在每个不同的 n_1，我们每次考虑不同的备择假设（在复合假设下），使得在不同的检验中，β 取某一常值. 因此，随着 n_1 的增加，α 和 β 固定不变，所考虑的备择假设随之变化.

上面所讲的可再用图 2-5 表现出来. 考虑参数 p 连续地趋近于 $p = 0.5$，随着 n_1 增大，图 2-5 中的 β 可以保持不变.

在备择假设下，对于每个 n_1，考虑计算有相同的 α 和 β 值的第二个检验的样本量 n_2 值. 那么对于原来检验序列中的每个检验，都有一列相对效率 n_2/n_1 值，若随着 n_1 增大，n_2/n_1 趋近于一个常数，且不随着 α 和 β 值的变化而改变，那么该常数称为第一个检验对第二个检验的渐近相对效率（asymptotic relative efficiency），或更准确些，是第一个检验序列对第二个检验序列而言的. 有时也称这样定义的渐近相对效率为 Pitman 效率（Pitman efficiency），以区分其他定义的渐近相对效率.

定义 2.4.4 令 n_1 和 n_2 分别是在相同的显著性水平下，有相同功效的两个检验 T_1 和 T_2 的样本容量. 如果 α 和 β 固定，当 n_1 趋于无穷时，极限 n_2/n_1 存在，且与 α 和 β 独立，那么，n_2/n_1 的极限称为第一个检验对第二个检验的渐近相对效率（A. R. E）.

在我们的问题中，为了寻找最大功效的检验，通常要找出具有最大渐近相对效

率的检验，因为功效依赖于太多因素．因此一个检验相对于另一个检验的 A. R. E 是很重要的.

通常两个检验的 A. R. E 计算起来比较困难．各种成对组合检验的 A. R. E 的全面研究本身就可以构成一本书的主题．Noether(1967a)写的书就涵盖了许多关于 A. R. E 的重要的研究结果．同时 Stuart(1954)与 Ruist(1955)对此也有进一步的研究.

所以 A. R. E 可以代替相对效率表．但是，如果样本无限（从而不可能），那么如何用 A. R. E 呢？对于小样本量的精确相对效率的研究表明，在很多实际应用中，A. R. E 可作为相对效率一个很好的近似．因此，A. R. E 简洁地概括了两个检验的相
112 对效率.

保守检验

在讨论一个检验时，我们有时还要考虑它的保守性（conservative）.

定义 2.4.5　一个检验称为是保守的，如果真实的显著性水平比规定的显著性水平小.

有时，计算一个检验的精确的显著性水平是很困难的，这时要使用近似计算 α 的一些方法，从而用近似值来作为显著性水平．如果近似的显著性水平比真实的显著性水平（但未知）大，则检验是保守的，并且我们知道犯第一类错误的风险没有规定的那么大.

习题

1. 一枚硬币掷 5 次．试验者记录每一次投掷的观测结果，将研究对象的眼睛蒙起来，猜测硬币落地时的“状态”，以检验他是否有超感知觉．零假设为研究对象猜对的概率是 $p=0.5$，而备择假设为 $p>0.5$．临界域为 5 次全部猜对.
   (a) 求 α.　　(b) 功效函数是什么？
   (c) 画出功效函数的图像.　　(d) 检验是无偏的吗？
2. 检验两种皮鞋中哪种更耐用．制造 8 双皮鞋，每双中除了其中一只由皮革 A 制做，另一只由皮革 B 制做外，这 8 双鞋一模一样．这些鞋正常使用一段时间后，再判断哪种皮革更耐用．令 X 表示判别得出 A 制成更耐用鞋的对数．零假设为 $p=0.5$，这里 p 为更耐用的鞋是由 A 制成的概率，而 H_1 为 $p\neq 0.5$，临界域为 $X=0,1,7,8$.
   (a) 求 α.　　(b) 功效函数是什么？
   (c) 画出功效函数的图像.　　(d) 检验是无偏的吗？
3. 令 $T_1,T_2\cdots$ 表示一检验序列，并假设 T_n 的功效 P_n 为 $P_n=n/(n+10)$．该检验序列是相合的吗？
113 4. 令 $T_1,T_2\cdots$ 表示一检验序列，并假设 T_n 的功效为 $n/(2n+10)$．该检验序列是相合的吗？
5. 对于假设检验 $H_0:p=1/2,H_1:p=3/4$，用显著性水平相等的两种检验 T_1 和 T_2 来检验．如果 T_1 的样本量是 20，那么 T_2 用 35 个样本时才能达到和 T_1 一样的功效.
   (a) T_2 相对于 T_1 的效率是多少？　　(b) T_1 相对于 T_2 的效率是多少？
6. 对于假设检验 $H_0:p=1/2,H_1:p\neq 1/2$，用显著性水平相等的两种检验来检验．当 T_1 的样

本量为 15 时，T_2 需要 30 个样本，它们的功效函数才能在特定备择假设 $p=1/3$ 时相等.

(a) T_2 相对于 T_1 的效率是多少?

(b) 备择假设为 $p=2/3$ 时，它们的效率一定相等吗?

思考题

1. 假设 T_2 相对于 T_1 的渐近效率是 0.75，且对于有限样本相对效率总是大于渐近相对效率. 如果试验者更愿意使用检验 T_2，且希望它的功效至少和样本量为 24 的检验 T_1 相等，那么检验 T_2 的最小样本量是多少?
2. 假设 T_1 相对于 T_2 的 A. R. E. 为 $3/\pi$，且 T_3 相对于 T_2 的 A. R. E 为 $2/\pi$. 那么 T_1 相对于 T_3 的 A. R. E 是多少?

2.5 非参数统计评述

本节我们试图区别术语*参数统计*（parametric statistics）与*非参数统计*（nonparametric statistics），尽管对于专业统计学家这些不同之处也不总是能区分得很清楚. 我们使用术语*非参数的*（nonparametric），更具体地说是*无分布的*（distribution-free），它们可以互相代替，尽管一些统计学家认为二者之间仍有差别. 我们就分析数据时什么时候使用非参数统计，什么时候参数法更有利给出一些指导.

使用优良方法

首先我们讨论假设检验和置信区间. 本章已经指出假设检验要基于一个好的统计量，它对零假设和备择假设间的差别应是敏感的，并且在零假设下它的概率分布已知. 置信区间是假设检验的逆推，因为置信区间是由数据不能拒绝的零假设的集 114
合. 所以，一个好的（有效的）假设检验对应于一个好的（短的）置信区间，反之亦然.

例如，样本均值 $\bar{X}$ 对于检验总体均值 μ 是一个好的检验统计量，因为它对总体均值的不同很敏感. 类似 S 和 s 对于推断总体标准差 σ 是好的检验统计量. 但是，$\bar{X}$,S 和 s 的概率分布取决于 X 的总体概率分布，但这通常是未知的.

参数方法

如果总体概率分布是正态分布，那么 $\bar{X}$，或一些基于 $\bar{X}$ 的统计量，可用于检验关于 μ 的假设，或求出估计 μ 的置信区间，因为零分布是已知的. 同样地，如果分布是正态分布，基于 S 或 s 和样本标准差，我们可以检验关于总体标准差 σ 的假设和构造 σ 的置信区间.

以上都称为*参数方法*（parametric method），因为它们都是在已知总体分布函数时有效. 任何假设检验或置信区间都是基于这样的假设：总体分布函数已知或只带有一些未知参数，这称为参数方法.

但是我们如何才能确定总体的概率分布是正态分布，或其他分布呢？答案很简单，我们不能. 在基于正态分布做假设检验之前，我们首先能通过考察数据来判断它是否来自于正态分布，或做一个考察数据非正态性的假设检验.

大多数参数方法都基于正态假设，因为检验背后的理论可以基于总体正态分布推出. 对于正态分布的数据，一些方法和结果是有效的. 其他的参数方法也是基于总体服从其他某一特定分布，如指数分布，威布尔分布等.

稳健方法

没有任何一个总体是服从精确的正态分布或其他任何已知的分布. 假如总体分布是近似正态的，那么通常（不总是）基于正态分布来使用这种方法是安全的. 但是，如果数据看起来显然来自非正态分布，或不适用于参数方法的分布，那么这时应当考虑非参数方法.

尽管一种分析数据的方法背后的某个假设条件不成立，但它是还近似有效的，

115 那么就认为这种方法对这一假设条件是*稳健的*（robust）. 一般说来，稳健一词是指基于正态假设的方法，而即使潜在的总体分布是非正态的，检验统计量也有近似相同的零分布.

一些参数检验，例如，一样本 t 检验或两样本 t 检验，特别是当样本量很大时，对于正态假设条件是稳健的，这就意味着检验统计量的零分布近似于正态总体所对应检验统计量的零分布，并且试验者可以将检验统计量的值与总体是正态时的精确的 t 分布表对应起来；即使总体是非正态的，我们也有信心认为表中的分位数，能够很好地近似检验统计量的真实的分位数.

然而，正因为方法是稳健的，所以不能确保当总体是非正态时，该方法一定像正态时那么有效. 因此使用一种统计方法我们不仅要问，它稳健吗？还要问，它有效吗？统计方法当然应该是稳健的，这样，使得到的显著性水平接近真实显著性水平. 但更应该是有效的，以便有效地利用和处理数据，以及拒绝错误的零假设.

非参数方法

非参数方法和参数方法都基于一些共同的假设，如假设样本是随机样本. 但是，非参数方法不假定特定的总体概率分布，因此对于来自任何未知概率分布总体的数据，它都适用.

非参数方法对总体分布假设是非常稳健的，因为它们对于所有的分布都同样有效.

如果总体分布函数比正态分布轻尾，例如均匀分布，那么基于正态假设的参数方法一般会得到好的功效，等于或大于在第5章将要介绍的基于秩的非参数方法得到的功效. 例如，意见调查数据是轻尾数据，问卷答案由1到5或1到7构成. 尽管答案的分布是离散的，且可能是不对称的，从而显然是非正态的. 但是由于它是轻尾的，所以在关于总体均值的假设检验中，从好的功效角度来讲，通常基于正态的

参数方法比非参数方法更受欢迎.

另一方面，如果总体分布函数比正态分布重尾，例如指数分布（第6章介绍），对数正态分布（数据的对数服从正态分布），卡方分布（属于伽马分布族），以及许多其他合理总体模型中出现的分布，那么基于正态假设的参数方法一般比基于秩的 116
非参数方法的功效要低.

包含离群值（outlier）的数据是来自重尾分布的典型数据，离群值的观测值比样本中其他的观测都大很多或小很多. 在这种情况下，考虑使用非参数方法来分析数据是非常重要的，例如第5章中所介绍的秩方法，因为秩方法的功效比基于正态假设的参数方法功效要高.

渐近分布自由

许多参数检验对于非正态假设条件是稳健的，也是渐近分布自由的（asymptotically distribution-free）. 这意味着随着样本容量的增加，方法变得更加稳健，对于无限样本容量的情形，方法是精确的且不依赖于总体分布. 通常，基于样本均值渐近无分布的参数方法的理论基础是中心极限定理. 在2.2中构造总体均值μ的渐近置信区间时，使用的就是上面的方法.

不应该只因为一种统计方法是非参数的，稳健的，或渐近无分布的，我们就更偏好它. 不论样本容量是多少，尽管方法是渐近分布自由的，参数检验的相对功效或置信区间的相对大小，和非参数方法比起来，通常有好有坏. 上面关于各种类型数据的统计方法偏好的讨论，不论样本容量的大小，都是适中的.

记住，我们考虑的绝大多数方法都是相合的，也就是说样本量的增大意味着绝对功效的变高. 如果样本量足够大，使得用一功效较小的检验也能够拒绝零假设，或用效率较低的方法得到的置信区间的长度，对试验者的要求来说已经足够短，那么仔细选择功效较大的方法就显得没有必要了. 当然，在选择分析数据的统计方法时，试验者还要考虑许多的其他方面.

名义数据的分析方法

正如我们引言所提到的，大多数人想到的非参数方法都是将在第5章和第6章介绍的基于秩的方法，因为秩方法是一些参数检验的合理选择，比如，对具有度量是区间或比率尺度的数据所进行的t检验和方差分析. 但是，非参数方法还可以用于采用名义或次序尺度的定性数据. 117

如果没处理过定性数据，至少是区间数据，那么很难想象名义或次序数据的总体概率分布的定义. 所以，没有参数方法适用于纯粹的名义或次序数据. 分析定性（名义）数据或仅仅已知次序或秩的数据，只能采用非参数方法. 第3章和第4章将介绍分析定性数据的方法，第5章和第6章的大多数方法更适合于分析次序数据.

非参数的定义

我们给出如下非参数的定义，它似乎比较有效.

定义 2.5.1 一种统计方法称为非参数的，如果它至少满足下面的法则之一：

1. 该方法适用于分析名义尺度数据.
2. 该方法适用于分析次序尺度数据.
3. 该方法适用于分析区间或比率尺度数据，这里除了有无限多个未知参数外，由随机变量分布函数所产生的数据，要么是非特定的，要么是特定的.

例2.3.1中检验的数据是名义数据（次品或非次品），所以由上面定义中的第一个法则，该检验是非参数的. 例2.3.1中检验的数据是次序数据，因此由第2条法则它也是非参数的. 几乎所有的非参数假设检验都满足这两条法则之一. 2.2节中的点估计满足第3个法则，并且在5.7,5.10和5.11节中基于对称分布的方法也满足第3个法则. 因此，我们认为它们是非参数的.

本书主要讨论的是假设检验和构造置信区间. 不幸的是，这经常给试验者一种错觉，他们如果不做假设检验或构造置信区间，就好像不是在做统计分析. 其他形式的统计分析也同样重要，例如总体的描述性，数据的解释，未知事件的预测以及点估计等.

这些其他形式的推断很大程度上依赖于试验者的经验和良好的判断力，而不是复杂的概率论证；所以我们认为它们太难，而不在本书中不加以讨论. 我们只把与假设检验和置信区间有关的一些复杂概率论证阐述清楚，以帮助那些已经具备足够
118 经验和良好判断力的试验者.

解决其他几类问题的已有非参数统计方法在本书中没有涉及，这些领域（和读者感兴趣的某些参考文献）包括生物鉴定（Miller,1973,Chmiel,1976），生存曲线（Susarla 和 Van Ryzin,1976,Tarone 和 Ware,1977）和纵向数据研究（Ghosh,Grizzle 和 Sen,1973）. Aitchison 和 Aitken(1976)，与 Bhapkar 和 Patterson(1977)讨论了多元方法，这也是 Puri 和 Sen(1971)书中所讨论的主题. 对于识别分析可参见 Gessaman 和 Gessaman (1972),Broffitt, Randles 和 Hogg(1976)，Randles，Broffitt，Ramberg 和 Hogg(1978),与 Conover 和 Iman(1980).

稳健方法在一定程度上依赖于总体分布函数，但对于偏离假设分布形式不是很敏感. 5.12节中简要地讨论了稳健方法. 更完整的讨论可在 Govindarajulu 和 Leslie (1972),Hogg(1974),Pearson 和 Please(1975),Policello 和 Hettmansperger(1976)以及5.12节所引入一些参考文献中找到. 非参数统计领域的概况介绍可见 Kendall 与 Sundrum(1953),Blum 与 Fattu(1954),Savage(1969),Bell(1964)和 Govindarajulu(1976)的相关文章或 Tate(1957),Fraser(1957),Walsh(1962),Noether(1967),Pierce(1970),Hollander 和 Wolfe (1973), Tapia (1978), Randles (1979), Buringer (1980), Henley (1981),Pratt(1981)以及 Manoukin(1986)的相关著作.

2.6 第2章复习题

1. 盒子里装有7张票，5张是学生票，另外2张是教师票. 从盒子中无放回地抽取2张票，来确定两名胜利者. 零假设是随机抽取，备择假设是作弊抽取，使得第一张票是教师票，第二张票从剩下的6票中随机抽取.
   (a) 假设判决准则是若所有的票都是教师的，则拒绝零假设. 求 α 和功效.
   (b) 假设判决准则是若第一张票是教师票，则拒绝零假设. 求 α 和功效.
   (c) 一些人偏爱 (a) 中的检验，因为它有较小的显著性水平. 另外一些人偏爱 (b) 中的检验，因为它有较大的功效. 讨论每种检验的社会影响. 你希望用哪种检验?
2. 下面各随机变量的度量尺度是什么?
   (a) 对某膳食，体重增加（或减少）的磅数.
   (b) 堪萨斯皇家队在职业棒球联赛中的排名.
   (c) 学生的学号.
   (d) 某篮球运动员的平均得分.
   (e) 奥运会比赛中某花样滑冰运动员的得分. 119
3. 两名学生通过下棋来比试. 法则为进行7局比赛，平局不算，如果一人至少赢了6局，那么就认为他们的棋力不在同一档次.
   (a) H_0 是什么?
   (b) H_1 是什么?
   (c) 写出样本空间中任意一点.
   (d) 列举临界域中所有的样本点.
   (e) 求出显著性水平.
   (f) H_0 是简单假设还是复合假设?
   (g) H_1 是简单假设还是复合假设?
   (h) 功效函数的方程式是什么?
   (i) 检验是无偏的吗?
   (j) 这里你会做出什么假定?
4. 客人离开后，餐馆经理将客人的支票放到收银机中，随后，审计员核查经理的工作，他发现有12个错误，其中10个错误对客人有利，而2个错误对经理有利. 令零假设为经理犯的错误对自己和客人有利的概率相等，而备择假设是经理犯的错误对自己和客人有利的可能性不等.
   设临界域为对客人有利的错误大于等于10个，或对经理有利的错误大于等于10个.
   (a) 求出该检验的显著性水平 α.
   (b) 如果经理犯错误对客人有利的概率是对自己有利概率的3倍，求检验的功效.
   (c) 画出功效曲线的图像. 该检验是无偏的吗? 加以解释说明.
5. 汽车生产流水线的最后阶段会有12%的车辆不能通过检查，需要加以特别处理. 设 X 为每4辆中不能通过检查的汽车数量，假设汽车能否通过检查是相互独立的.
   (a) 求 X 的均值和标准差.
   (b) 求 X 的中位数和四分位极差.
6. 继续题5，生产6组汽车，每组4辆，在生产线最后阶段每组车中没通过检查的数量分别为0,0,0,1,1,2.
   (a) 求样本均值和标准差.
   (b) 求样本中位数和四分位极差.
   (c) 由数据估计一辆汽车不能通过检查的概率.

7. 讨论习题5中的度量尺度.

120 (a) 每辆车的度量尺度是什么? (b) X 的度量尺度是什么?

8. 得克萨斯州的所有家庭中，20%没有车，30%有1辆车，30%有2辆车，10%有3辆车，剩下的超过3辆车. 随机抽取10个家庭，结果如下：没有车的家庭是3个，有1辆车的家庭2个，有2辆车的家庭2个，有3辆车的家庭1个，有4辆车的家庭1个，有5辆车的家庭1个.
   (a) 求每个家庭车辆拥有量的总体中位数. (b) 求总体四分位极差.
   (c) 讨论在这个问题中为什么用总体中位数比总体均值要好.
   (d) 求每个家庭拥有车数的样本中位数. (e) 求车数的样本均值.
   (f) 求样本标准差，用 S 而不是 s. (g) 画出总体分布函数图像.
   (h) 画出经验分布函数图像.

9. 令 X 为上下班高峰期城市快速路上任一辆机动车中的人数. X 的概率为 $P(X=1)=0.40$，$P(X=2)=0.30$，$P(X=3)=0.20$ 及 $P(X=4)=0.10$. X 的一组随机样本中有10个观测值，如下所示：

   4, 1, 1, 2, 1, 2, 3, 1, 4, 1

   求样本均值. 画出经验分布函数图像（见定义2.2.1）. 求样本四分位极差. 将这些值分别与总体均值，分布函数（见定义1.3.4）和总体四分位极差进行比较.

10. 若零假设成立，那么每位顾客选择红盒麦片或蓝盒麦片的可能性相等. 若备择假设成立，顾客喜欢挑选蓝盒与红盒的比例为3比1. 假设每位顾客挑选盒装麦片是独立的. 如果最初的20个顾客中不少于15个人选择蓝盒，那么拒绝零假设. 求显著性水平，求功效. 如果实际上观测最初的20个顾客中有17个人选择蓝盒，求 p 值.

11. 7个男生应征3个夏令营辅导员职位. 根据身高将学生由高到低排序，分别用1（最高）到7（最低）表示. 零假设为每名学生被选中的概率相等，备择假设是3名身高较高的学生被选中的可能性是4名身高较矮学生的2倍. 假设每个学生是否被选中相互独立. 检验统计量是被选中的3名学生身高排序的秩和，判决法则是若检验统计量小于等于6，则拒绝零假设.
   (a) 零假设是简单假设还是复合假设?
   (b) 求显著性水平. (c) 求检验的功效.

121 12. 一种基因理论认为某两只狗的后代为斑点狗的概率为25%，令此为零假设. 另一理论则认为每只小狗有斑点的概率为75%，令此为备择假设. 一窝小狗出生了，8只中有5只是带斑点的小狗.
   (a) 用目标显著性水平0.05来求出一保守检验的临界域.
   (b) 在你的检验中，精确显著性水平值是多少?（采用精确的公式，正态逼近计算得到两个稍微不同的结果）
   (c) 在你的检验中，精确的功效是多少?（使用表格计算出答案）
   (d) 这种情况下，p-值是多少?
   (e) 该检验是无偏的吗? 请解释.
   (f) 备择假设是简单假设还是复合假设?

13. 掷一枚非均匀硬币“正面”出现的概率为2/3，“反面”出现的概率为1/3. 该硬币掷10

次，结果出现5次“正面”，5次“反面”. 设 X 为“正面”出现的次数.

(a) 画出 X 的总体分布函数.　　(b) 求出总体中位数.

(c) 画出 X 的经验分布函数.　　(d) 求出样本中位数.

14. 一个学生做3道多项选择题，每道题有5个可能答案，如果他学过该课程，就有超过80%的可能性答对每道题；如果他没学过该课程（零假设），则每道题他选择任一答案的概率相等. 若他3道题全部答对，拒绝零假设.

(a) 求出显著性水平.　　(b) 求出检验的功效. 122

(c) 还要做出哪些题目中没有明确提出的假定?

第3章 基于二项分布的检验[1]

导 言

在第1章中，我们介绍了用二项分布来描述 n 次掷硬币试验中正面出现次数的概率. 在更一般的形式为 n 次独立基本试验中的每一次结果，或以概率为 p "成功"，或以概率 $q=1-p$ "失败". 二项分布描述了 n 次试验中恰有 k 次成功的概率. 表 A3 给出了一些二项分布函数.

应用科学中的许多试验都可以用这种方法来建模. 例如，一些顾客到达商店，自主地决定买或不买某种商品；给动物用某种药物，它们治愈或没有治愈. 这种例子几乎可以在任何领域中找到，我们就可以用一些熟知的基于二项分布的，最简单的统计方法来分析这些情况下所获得的数据. 本章我们将给出几个可行的方法，而另外章节将围绕基于二项分布的其他方法展开. 学习了本章所给的各种检验后，读者应当能够变通所学的方法，使之适用于所给的试验情形.
123

3.1 二项检验与 p 的估计

我们已经给出过二项检验的一个例子，例 2.3.1 把二项检验用于质量控制问题. 本章则介绍比例 2.3.1 更多的内容，说明简单二项检验的多种用法和奇妙变化. 只要灵活运用得好，二项检验可用来检验几乎所有的假设和所有类型的统计数据分析. 在有些场合，二项检验是最有效的检验，这时检验是用参数和非参数统计来要求的，而在另外一些场合，二项检验是比较有效的，我们只能用非参数统计来要求. 然而，即使是在比较有效的情形下，人们也更愿意选用二项检验，因为它操作简单，易于解释，有时它有足够的有效性，使得在零假设应该拒绝时足以拒绝原假设.

我们现在正式介绍二项检验，并同时介绍二项检验的格式. 为了使读者方便和非参数方法使用者容易掌握，我们觉得叙述一下检验的格式是有必要的.

▶二项检验

数据 样本中包含 n 次独立基本试验的结果，每个结果或者是"类1"或者是"类2"，但两类不能同时出现. 类1的观测数是 O_1，类2的观测数是 $O_2=n-O_1$.

---

1. 第3章的复习题包括在第4章的复习题中.

假定条件

1. n 次基本试验相互独立.

2. 每次基本试验中结果“类1”出现的概率是 p，且对所有 n 次基本试验有相同的 p.

检验统计量 由于我们关注的是结果“类1”出现的概率，我们令检验统计量 T 是结果为“类1”的次数，即

$$T = O_1 \tag{1}$$

零分布

令 p^* 是零假设中给定的概率. T 的零分布是参数为 $p = p^*, n =$ 样本容量的二项分布. 对于 $n \leqslant 20$ 和选定的 p 值，表 A3 中列出了 T 的零分布. 124

对其他的 n 值和 p 值，我们可以用正态分布近似，即 T 的 q 分位数 x_q 可以由下式近似给出

$$x_q = n \cdot p + z_q \sqrt{n \cdot p \cdot (1-p)} \tag{2}$$

其中，z_q 是标准正态随机变量的 q 分位数，见表 A1.

假设 令 p^* 是某个给定的概率，$0 \leqslant p^* \leqslant 1$，假设可以是下列 3 种形式之一.

A.（双边检验）

$$H_0: p = p^*$$
$$H_1: p \neq p^*$$

理想水平 α 的拒绝域对应于 T 零分布的两边，其中左边水平为 α_1，它近似于$\alpha/2$，右边水平为 α_2，也是近似于 $\alpha/2$，其真实的显著水平是 $\alpha_1 + \alpha_2$，由于 T 的离散性，这一真实显著水平很少为 α.

因此，对给定的特殊 p^* 和 n 的值，我们从表 A3 中找到 t_1，使得

$$P(Y \leqslant t_1) = \alpha_1 \tag{3}$$

并找到 t_2，使得

$$P(Y \leqslant t_2) = 1 - \alpha_2 \tag{4}$$

其中，Y 是参数为 p^* 和 n 的二项随机变量.

如果 $n > 20$，我们用正态分布逼近，即用（2）式去近似 t_1, t_2，其中 t_1, t_2 分别是参数为 p^* 和 n 的二项随机变量的 $\alpha/2$ 分位数和（$1 - \alpha/2$）分位数，只要在（2）式中分别令 $q = \alpha/2$ 和 $q = 1 - \alpha/2$ 即可.

如果 $T \leqslant t_1$ 或 $T \geqslant t_2$，则拒绝 H_0，否则接受零假设.

p-值（尾概率）是两概率 P（Y 小于或等于观测值 T）和 P（Y 大于或等于观测值 T）中较小的一个的2倍，对于 $n \leqslant 20, p = p^*$，p-值可以从表 A3 中获得；对于 $n > 20$，可利用表 A1 和如下近似公式获得 125

$$P(Y \leqslant t_{\text{obs}}) \cong P\left(Z \leqslant \frac{t_{\text{obs}} - n \cdot p^* + 0.5}{\sqrt{n \cdot p^*(1-p^*)}}\right) \tag{5}$$

和

$$P(Y \geq t_{\text{obs}}) \cong 1 - P\left(Z \leq \frac{t_{\text{obs}} - n \cdot p^* - 0.5}{\sqrt{n \cdot p^*(1-p^*)}}\right) \tag{6}$$

其中，引入0.5是作为改进二项分布正态逼近的一种“连续性修正”.

B.（左单边检验）

$$H_0: p \geq p^*$$
$$H_1: p < p^*$$

由于小的T值预示着H_0是假的，于是水平为α的拒绝域是$\{T: T \leq t\}$，其中t由表A3获得，参数取为p^*和n. 所以

$$P(Y \leq t) = \alpha \tag{7}$$

其中，Y是参数为p^*和n的二项随机变量.

如果$n > 20$，我们就用正态逼近，即用（2）式去近似t，其中t是以参数为p^*和n的二项随机变量的α分位点，只要在（2）式中令$q = \alpha$即可.

如果$T \leq t$，则拒绝H_0，否则接受零假设.

p-值是概率P（Y小于或等于观测值T），当$n \leq 20, p = p^*$时，它可从表A3中获得，如果$n > 20$，可利用表A1和如下近似公式获得

$$P(Y \leq t_{\text{obs}}) \cong P\left(Z \leq \frac{t_{\text{obs}} - n \cdot p^* + 0.5}{\sqrt{n \cdot p^*(1-p^*)}}\right) \tag{8}$$

其中，引入0.5是作为改进二项分布正态逼近的一种“连续性修正”.

C.（右单边检验）

$$H_0: p \leq p^*$$
126
$$H_1: p > p^*$$

因为大的T值预示着H_0是假的，于是水平为α的拒绝域是$\{T: T \geq t\}$，其中t由表A3获得，参数取为p^*和n. 所以

$$P(Y \leq t) = 1 - \alpha \tag{9}$$

其中，Y是参数为p^*和n的二项随机变量.

如果$n > 20$，我们就用正态逼近，即用（2）式去近似t，其中t是以参数为p^*和n的二项随机变量的$1-\alpha$分位点，只要在（2）式中令$q = 1-\alpha$即可.

如果$T > t$，则拒绝H_0，否则接受零假设.

p-值是概率P（Y大于或等于观测值T），当$n \leq 20, p = p^*$时，它可从表A3中获得，如果$n > 20$，可利用表A1和如下近似公式获得

$$P(Y \geq t_{\text{obs}}) \cong 1 - P\left(Z \leq \frac{t_{\text{obs}} - n \cdot p^* - 0.5}{\sqrt{n \cdot p^*(1-p^*)}}\right) \tag{10}$$

其中，引入0.5是作为改进二项分布正态逼近的一种“连续性修正”.

计算机辅助　一些计算机软件包可以进行这种检验并给出p-值，包括*Minitab*，*S-Plus*和*StatXact*. *Minitab*也可以计算出功效，以及达到要求功效水平而所需的样

本容量. ◀

例 3.1.1

据估计，目前做前列腺癌手术的男性中有一半正遭受某种副作用的影响. 为了努力减轻这种副作用的可能性，FDA 研究了一种新的手术方法. 19 例受手术者只有 3 人有这种不良副作用，由此得出这项新手术方法能有效减轻副作用，这个结论会可靠吗?

令 p 为患者遭受副作用影响的概率，这是一个左单边检验，$H_0: p \geq 0.5$ 对 $H_1: p < 0.5$. 如果目标值 α 是 0.05，那么拒绝域是 $\{T: T \leq 5\}$，而实际上 $\alpha = 0.0318$（见表 A3, $n=19, p=0.5$).

观测值 T 为 3，所以拒绝 H_0，得出新方法在降低副作用可能性方面有效，其 p-值为

$$P(T \leq 3) = 0.0022$$

它很小，表示样本数据强烈地拒绝零假设. ■ 127

当精确的方法可行时，人们总是用精确的方法，但为了解释正态逼近如何好，我们考虑例 3.1.1. 从（2）式中可以得到近似的 0.05 分位点，

$$x_{0.05} = 19(0.5) + (-1.6449)\sqrt{19(0.5)(0.5)} = 5.9$$

得到了与前面相同的拒绝域. 从（5）式我们可获得精确 α 值的估计

$$P(Y \leq 5) \cong P\left(Z \leq \frac{5 - 19(0.5) + 0.5}{\sqrt{19(0.5)(0.5)}}\right) = 0.033$$

它很接近于 α 的精确值 0.032. 精确的 p-值也可以由（5）式估计

$$P(Y \leq 3) \cong P\left(Z \leq \frac{3 - 19(0.5) + 0.5}{\sqrt{19(0.5)(0.5)}}\right) = 0.003$$

同样，它与真实的 p-值 0.002 也很接近.

例 3.1.2

在简单的孟德尔遗传试验中，将两种特殊基因类型的植物进行杂交，产生的后代中，可能有 1/4 是“矮”型，3/4 是“高”型的. 在一项验证某条件下简单孟德尔遗传假设是否成立的试验中，杂交后代中有 243 个矮植物和 682 个高植物. 如果“类 1”代表“高”，那么 $p^* = 3/4$，则 T 等于高植物的个数. 孟德尔遗传规律的零假设等价于假设模型，即

$$H_0: p = 3/4$$

感兴趣的备择假设是双边的，即

$$H_1: p \neq 3/4$$

因为 $n = 243 + 682 = 925$，那么水平 $\alpha = 0.05$ 的拒绝域可以通过（2）式所给的大样本逼近得到. 所以，拒绝域为 $\{T: T \leq t_1\} \cup \{T: T > t_2\}$，其中

$$t_1 = np^* + z_{0.025}\sqrt{np^*(1-p^*)}$$

128

$$= (925)(\tfrac{3}{4}) + (1.960)\sqrt{(925)(\tfrac{3}{4})(\tfrac{1}{4})}$$

$$= 667.94 \tag{11}$$

$$t_2 = np^* + z_{0.975}\sqrt{np^*(1-p^*)}$$

$$= (925)(\tfrac{3}{4}) + (1.960)\sqrt{(925)(\tfrac{3}{4})(\tfrac{1}{4})} = 719.56 \tag{12}$$

此试验中的 T 值为 682，所以接受零假设.

p-值可由（5）式计算得出

$$P(Y \leqslant 682) \cong P\left(Z \leqslant \frac{682 - 693.75 + 0.5}{13.17}\right) = P(Z \leqslant -0.8542) = 0.196 \tag{13}$$

其中，Z 是标准正态分布，其概率可在表 A1 中查到，这个单边检验 p-值的 2 倍就是双边检验的 p-值 0.392.

显著水平至少在 0.392 时才可能拒绝 H_0，所以数据与零假设吻合得较好. ■

前面的例子解释了双边形式的二项检验，单边二项检验在例 2.3.1 中已经给了解释.

□**理论** 通过比较二项检验中的假设，以及例 1.3.5 和例 1.2.8 中的假设，我们很容易看出，二项检验中的检验统计量是二项分布的，即如果 T 等于基本试验结果中“类 1”的个数，其中基本试验是相互独立的，且每次基本试验得“类 1”结果的概率为 p（如假设中所陈述），那么 T 服从参数为 p，n 的二项分布. 在零假设成立时，拒绝域的大小在 p 等于 p^* 时达到最大. 所以对于参数 n 和 p^*，表 A3 可用来确定 α 的精确值. □

正如前面提到的，假设检验只是统计推断中的一个分支. 现在我们来讨论另外一个分支，即区间估计（interval estimation）. 如果我们想对某个总体的一个未知参数做出某些推断，合理的做法是抽查这个总体中的一个随机样本，并且基于这个样本得出有关这个总体参数的一些论断，这种推断可能是“总体参数在 a 和 b 之间”，其中 a 和 b 是由样本得到的两个实数. 由于 a 和 b 是由样本值计算得出的，因而是两个统计量的实现值. 这两个统计量提供了区间的左端点和右端点，我们分别用 L 和 U 表示，代表“左”和“右”. 从 L 到 U 的区间称为区间估计量（interval estimator）. 总体未知参数落在此区间内的概率称为置信系数（confidence coefficient）. 区间估计量和置信系数给
129 我们提供了置信区间（confidence interval）.

对一个特定事件发生的概率 p 未知，其寻找 p 的置信区间方法与二项检验密切相关.

▶概率或总体比例的置信区间

数据 察看含有 n 个独立基本试验观测值的样本，并记 Y 为指定事件发生的次数.

假定条件

1. n 次基本试验互相独立.

2. 从一个基本试验到另一个基本试验，指定事件发生的概率 p 是常数.

方法 A　对于 $n \leqslant 30$，利用表 A4，置信系数是 0.90, 0.95 或 0.99. 只须给出样本值 n 和观测值 Y，我们利用该表，在对应栏里的交叉处，给出了所需置信区间的左、右限.

方法 B　对于 $n > 30$，或置信系数没有在表 A4 中列出的，用下列正态分布逼近

$$L = \frac{Y}{n} - z_{1-\alpha/2}\sqrt{Y(n-Y)/n^3} \tag{14}$$

和

$$U = \frac{Y}{n} + z_{1-\alpha/2}\sqrt{Y(n-Y)/n^3} \tag{15}$$

其中，$z_{1-\alpha/2}$ 是正态随机变量的分位数，它可从表 A1 中查出，其置信系数近似于 $1-\alpha$.

计算机辅助　计算机软件包可以算出二项参数 p（或总体比例 p）的置信区间，这些软件包包括 *Minitab*, *S-plus* 以及 *StatXact*. ◀

为了表达更清楚，在下面例子中将使用两种方法来计算置信区间.

例 3.1.3

在某个州随机选择 20 所高中，来检查它们是否达到国家教委提出的优秀标准. 调查发现有 7 所学校达到优秀，并且因此被评为“优秀”，那么该州所有高中符合评 130
为“优秀”比例 p 的 95% 置信区间是什么?

首先，我们假设该州高中的数量足够多，使得高中被评为“优秀”和“不优秀”是相互独立的.

因为我们假设抽取是随机的，那么对于所有学校 p 是相同的，它代表一个随机被抽到的学校被评为“优秀”的概率.

因为 $n=20, Y=7$，我们可以利用表 A4，由表 A4 给出的精确 95% 置信区间是 [0.154, 0.592].

方法 B，用基于中心极限定理的正态分布逼近，可得：

$$\begin{aligned} L &= \frac{Y}{n} - z_{0.975}\sqrt{Y(n-Y)/n^3} = 0.35 - (1.960)\sqrt{(7)(13)/(20)^3} \\ &= 0.35 - 0.209 = 0.141 \end{aligned} \tag{16}$$

和

$$U = 0.35 + 0.209 = 0.559 \tag{17}$$

由正态分布逼近得到的置信区间是 [0.141, 0.559]，它接近于精确区间，但是仍能看出它们的差距，这表明用精确置信区间的好处是显然的. ■

□**理论**　对于上面介绍的精确方法 A，如果用双边二项检验，置信区间包括所有 p^* 值，使得从样本中获得的数据能够接受

$$H_0: p = p^*$$

更确切地说，如果我们想形成一个 $(1-\alpha)$ 的置信区间，就需要观察样本并确定 Y，那么我们要问“对于给定的 Y，我们用什么 p^* 值，对于假设

$$H_0: p = p^*$$

使得一个双边二项检验（α 水平）可以接受 H_0?”，即这些 p^* 值应当在我们的置信区间
131 中，而使拒绝 H_0 的 p^* 值应当不在置信区间中. 由于二项检验的每一边有概率 $\alpha/2$，对给定的 Y 值，譬如说它是 y 或更大的值，用仅产生拒绝 H_0 的 p^* 来作为 L 的选取，则 p^* 的选择应满足

$$P(Y \geqslant y \mid p = p_1^*) = \frac{\alpha}{2} = \sum_{i=y}^{n} \binom{n}{i} (p_1^*)^i (1 - p_1^*)^{n-i} \tag{18}$$

所以 $L = p_1^*$. 然后对同样的 y 值，另一个 p^* 的选择应使得仅产生拒绝域的左边，即 p_2^* 满足

$$P(Y \leqslant y \mid p = p_2^*) = \frac{\alpha}{2} = \sum_{i=0}^{y} \binom{n}{i} (p_2^*)^i (1 - p_2^*)^{n-i} \tag{19}$$

令 $U = p_2^*$，我们知道（18）式和（19）式不可能用代数求解，只能通过搜索程序在计算机上求解而得到表 A4.

关于二项参数 p 的置信区间的更多内容，可参见 Clopper 和 Pearson(1934).

对于 L 和 U 的大样本逼近，可通过例 1.5.6 来获得，即是说：如果 Y 是一个二项随机变量，具有参数为 p 和较大 n，那么

$$Z = \frac{Y - np}{\sqrt{npq}} \tag{20}$$

是一个近似于标准正态分布的随机变量. 那么，如果 $z_{1-\alpha/2}$ 是表 A1 中（$1-\alpha/2$）分位数，并注意到 $z_{\alpha/2} = -z_{1-\alpha/2}$，故有

$$\begin{aligned} 1 - \alpha &= P\left(-z_{1-\alpha/2} < \frac{Y - np}{\sqrt{npq}} < z_{1-\alpha/2}\right) \\ &= P(-z_{1-\alpha/2}\sqrt{npq} < Y - np < z_{1-\alpha/2}\sqrt{npq}) \end{aligned}$$

对求概率中的不等式两边乘以（-1），不等式改变方向，

$$1 - \alpha = P(z_{1-\alpha/2}\sqrt{npq} > np - Y > -z_{1-\alpha/2}\sqrt{npq})$$

调换顺序，得

$$\begin{aligned} 1 - \alpha &= P(-z_{1-\alpha/2}\sqrt{npq} < np - Y < z_{1-\alpha/2}\sqrt{npq}) \\ &= P(Y - z_{1-\alpha/2}\sqrt{npq} < np < Y + z_{1-\alpha/2}\sqrt{npq}) \end{aligned}$$

再除以 n，得

132

$$1 - \alpha = P\left(\frac{Y}{n} - z_{1-\alpha/2}\sqrt{\frac{pq}{n}} < p < \frac{Y}{n} + z_{1-\alpha/2}\sqrt{\frac{pq}{n}}\right) \tag{21}$$

用更进一步地近似，在（21）式的根号中用估计量 Y/n 来估计 p，得到

$$\begin{aligned} 1 - \alpha &\cong P\left(\frac{Y}{n} - z_{1-\alpha/2}\sqrt{\frac{Y}{n}\left(1 - \frac{Y}{n}\right)\Big/ n} < p < \frac{Y}{n} + z_{1-\alpha/2}\sqrt{\frac{Y}{n}\left(1 - \frac{Y}{n}\right)\Big/ n}\right) \\ &\cong P(L < p < U) \end{aligned} \tag{22}$$

其中，L 和 U 与（14）式和（15）式中的相同. 这后面用 Y/n 对 p 的近似，其结果与置信区间和假设检验略有些不同，当样本量较大时，两者都可以用.

在上述过程中，对 L, U 同乘以样本容量 n，这样 nL 和 nU 就给出了 nP 的置信上、下限，它可用来检验包括二项随机变量均值在内的假设，因为

$$H_0: p = p^*$$

等价于

$$H_0: np = np^*$$ □

其他给出二项分布置信限的方法可参见 Anderson 和 Burstein(1967,1968). Quesenberry 和 Hurst(1964)及 Goodman(1965)则给出了处理多项比例的联合置信区间的方法.

习题

下面的每一个练习中，在需要的地方应清楚地陈述 H_0, H_1, T；判定原则；α；判定结果；p-值以及所用检验的名称.

1. 已知某种昆虫的 20% 显示出特性 A，在非正常的环境下得到 18 只这种昆虫，没有一只具有特性 A. 那么假设在这种环境下，此种昆虫和通常环境一样有 0.2 的概率显示特性 A，这合理吗？用双边检验.
2. 在一次安全月活动中，所检查的 16 辆车中有 6 辆车是不安全的. 检验零假设：这些车中有不多于 10% 的车是不安全的.（这个应用中哪个假设更可能是假的?）
3. 掷一对骰子 180 次，事件“两个点数之和为 7”共发生 38 次. 如果骰子是均匀的，那么出现“7”的概率是 1/6. 如果是不均匀的，那么概率更高.
   (a) 如果骰子是均匀的，用单边检验此次游戏出现“7”的次数是否正常.
   (b) 应用大样本逼近，求 P(出现“7”)的 95% 置信区间. 133
4. 在习题 2 中，不安全车真实比例的 90% 置信区间是什么？
5. 从服从未知分布 $F(x)$ 的随机变量 X 中得到如下独立的 20 个观测值.

142	134	98	119	131
103	154	122	93	137
86	119	161	144	158
165	81	117	128	103

   求 $F(100)$ 的 95% 置信区间.
6. 一个市民小组向市政府报告说，至少有 60% 的居民认同特殊的发行债券. 市政府随后就随机调查了 100 个居民，问他们是否认同这种特殊的债券，48 个人表示同意. 问这个市民小组的报告是否合理？
7. 最近的 20 个公司兼并案中，有 5 个因被兼并的公司的反对而流产. 假设它们是独立的事件，试估计一个兼并尝试被成功拒绝的概率，即，找到一个 95% 的置信区间.
   (a) 用表 A4.　　(b) 用表 A1.
8. 一个老师想调整一门继续教育课程的难度水平，来满足学生的需要. 他教了几次课，并且每次给学生们一个简单的评价调查问卷，他发现 12 个学生认为课程太简单，84 个学生认为课程合适，3 个人认为课程太难. 试检验零假设：学生们等可能地认为课程太简单或太难，双边备择假设：课程的难度水平需要改变，显著水平取 5%.

9. 设在20个独立基本试验观测中有3个成功，且检验的零假设是P(成功)$\leqslant 0.3$，备择假设是P(成功)>0.3，试用二项公式求精确p-值（临界值）.
10. 20个得克萨斯理工大学法律学院的毕业生参加了法律毕业考试，18个通过. 如果它表示整个得克萨斯理工大学法律学院毕业生的一个随机样本，问这能否证明整个得克萨斯理工大学法律学院毕业生通过法律考试的概率比州平均水平（70%）高？
11. 在水下战争演习中发射了20枚鱼雷，有15枚击中目标，求一枚鱼雷击中目标概率的90%置信区间.
    (a) 用表A4求解. (b) 用大样本逼近求解.
    (c) 讨论在求解问题中你所做的假设.
12. 70种化学检测试剂一起放在气体室里一段固定时间，往气体室里充入一定量的致命气体. 56种试剂对这种致命气体反应呈阳性，其他14种则没有呈阳性. 求在这种条件下能呈阳性概率的90%置信区间.

134

思考题

1. 连续性修正. 显然，如果Y服从二项分布，那么
$$P(Y\leqslant 4)=P(Y\leqslant 4.1)=\cdots=P(Y\leqslant 4.999)$$
因为Y只取整数值，如4,5等，相邻整数之间没有其他值，所以，用正态分布逼近二项分布时，我们应用哪个数：4,4.1，还是其他什么数？连续性修正（由于我们试图用一个连续分布，如正态分布去逼近一个离散分布，如二项分布）就是用离散分布的两个邻近值的中间数. 即，在二项分布中估计，我们用
$$P(Y\leqslant 4)\cong P\left(Z\leqslant\frac{4+0.5-np}{\sqrt{nqp}}\right)$$
估计$P(Y\leqslant 4)$，其中，Z服从正态分布，而4.5是4和5的中间数.
通常，连续性修正用在正态分布逼近二项分布时效果很好.
(a) 对于$n=20, p=0.1$，用表A3求$P(Y\leqslant 1)$的准确值，并用正态分布逼近来估计$P(Y\leqslant 1)$. 先不用连续性修正，再用连续性修正，看哪个估计更接近准确值？
(b) 重复(a)，而将$p=0.1$换为$p=0.3$，哪个估计更接近准确值？
2. 令Y_1, Y_2是两个独立的分别服从参数n_1, p_1和n_2, p_2的二项分布的随机变量.
(a) 证明$Y_1/n_1-Y_2/n_2$的均值为p_1-p_2.
(b) 证明$Y_1/n_1-Y_2/n_2$的方差是$p_1(1-p_1)/n_1+p_2(1-p_2)/n_2$.
(c) 说明可用$Y_1(n_1-Y_1)/n_1^3+Y_2(n_2-Y_2)/n_2^3$来估计$Y_1/n_1-Y_2/n_2$的方差.
(d) 如果$Y_1/n_1-Y_2/n_2$近似服从正态分布，证明（p_1-p_2）的一个置信度近似为$1-\alpha$的置信区间可由下式给出：
$$\frac{Y_1}{n_1}-\frac{Y_2}{n_2}-z_{1-\alpha/2}s<p_1-p_2<\frac{Y_1}{n_1}-\frac{Y_2}{n_2}+z_{1-\alpha/2}s$$
其中
$$s=\sqrt{Y_1(n_1-Y)/n_1^3+Y_2(n_2-Y_2)/n_2^3}$$
$z_{1-\alpha/2}$由表A1得到.

3.2 分位数检验和χ_p的估计

二项检验可以用来检验有关随机变量分位数的假设，此时，我们称之为分位数检验. 例如，我们检验一个随机变量 X 的随机样本值，看它的中位数是否为 17. 如果 X 的中位数是 17，那么应当大约各有一半的观测值落在 17 的两边，正如 $p=1/2$ 的二项分布那样. 如果有很少的样本观测值小于 17，那么 X 的中位数应当大于 17，如果有很多样本观测值小于 17，那么 X 的中位数小于 17. 135

度量尺度对于分位数检验至少是次序尺度，虽然二项检验只需要弱名义尺度来度量. 这是因为分位数几乎与度量的名义尺度没有关系. 如果被检验的随机变量是连续的，检验的假设是：

$$H_0: X\text{的}p^*\text{分位数是指定的}x^*$$

由分位数的定义，这等价于

$$H_0: P(X\leqslant x^*)=p^*$$

如果我们用 p 代表未知的概率 $P(X\leqslant x^*)$，则 H_0 变为：

$$H_0: p=p^*$$

这与二项检验的原假设是相同的. 检验统计量等于样本值小于或等于 x^* 的个数，可以用双边二项检验.

如果假设随机变量不是连续的，那么情况就没有这么简单了. 此时零假设为：

$$H_0: X\text{的}p^*\text{分位数是}x^*$$

等价于

$$H_0: P(X\leqslant x^*)\geqslant p^* \text{ 和 } P(X<x^*)\leqslant p^*$$

现在可以用二项检验，但是对这个假设检验的修改需要一些技巧，所以我们将给出单独检验的方法.

▶分位数检验

数据 令 $X_1,X_2,\cdots,X_n$ 是一组随机样本，数据由 X_i 的观测值组成. 136

假定条件

1. 这些 X_i 是随机样本（即，它们是独立同分布的随机变量）.
2. 度量尺度至少是须序的.

检验统计量 在这个检验中我们将用两个检验统计量. 令 T_1 等于观测值中小于等于 x^* 的个数，T_2 等于观测值中小于 x^* 的个数. 那么，当数据中没有数严格等于 x^* 的数时，$T_1=T_2$，否则，T_1 大于 T_2.

零分布 检验统计量 T_1 和 T_2 的零分布是二项分布，参数 n = 样本量，$p=p^*$ 和零假设一样. 在表 A3 中给出了 $n\leqslant 20$，和选定 p 值时的零分布.

对于其他 n，p 值，用正态分布逼近．即，T 的近似分位数 x_q 为

$$x_q = n \cdot p + z_q \sqrt{n \cdot p \cdot (1-p)} \tag{1}$$

其中，z_q 是标准正态随机变量的 q 分位数，在表 A1 中给出．

假设 令 x^*, p^* 为指定的值，$0 < p^* < 1$，则假设可能是如下三种形式中的一种．

A.（双边检验）

H_0: 第p^*个总体的分位数为x^*

[这等价于 $H_0: P(x \leqslant x^*) \geqslant p^*$ 和 $P(X < x^*) \leqslant p^*$，其中 X 与样本中 X_i 有相同的分布.]

H_1: x^*不是第p^*个总体的分位数

拒绝域对应于 T_2 其值太大［说明可能 $P(X < x^*)$ 大于 p^*］或对应于 T_1 其值太小［说明可能 $P(X \leqslant x^*)$ 小于 p^*］．和双边二项检验一样，通过表 A3，样本量 n，假设概率 p^*，可以得到拒绝域．找到 t_1，使得

137 $$P(Y \leqslant t_1) = \alpha_1 \tag{2}$$

其中，Y 服从参数为 n 和 p^* 的二项分布，α_1 是给定显著性水平的一半．得到 t_2，使得

$$P(Y \leqslant t_2) = 1 - \alpha_2 \tag{3}$$

其中，选 α_2 使得 $\alpha_1 + \alpha_2$ 大约等于给定的显著性水平．如果 T_1 小于等于 t_1，或 T_2 大于 t_2，拒绝 H_0，否则接受 H_0，显著性水平等于 $\alpha_1 + \alpha_2$.

对于 $n > 20$ 或表 A3 中没有的 p^* 值，由（1）式分别令 $q = \alpha/2$ 和 $q = 1 - \alpha/2$，求出 $t_1 = x_{\alpha/2}$ 和 $t_2 = x_{1-\alpha/2}$.

p-值是二项随机变量 Y 小于等于观测值 T_1，或大于等于 T_2 的概率中较小值的 2 倍，当 $n \leqslant 20, p = p^*$ 时，可以从表 A3 查出，对于 $n > 20$ 用表 A1，用

$$P(Y \leqslant T_1) \cong P\left(Z \leqslant \frac{T_1 - n \cdot p^* + 0.5}{\sqrt{n \cdot p^* (1-p^*)}}\right) \tag{4}$$

和

$$\mathrm{P}(\mathrm{Y} \geqslant \mathrm{T}_2) \cong 1 - \mathrm{P}\left(\mathrm{Z} \leqslant \frac{T_2 - n \cdot p^* - 0.5}{\sqrt{n \cdot p^* (1-p^*)}}\right) \tag{5}$$

两式与 0.5 作为“对连续性的修正”，来改进正态对二项分布的逼近.

B.（左边检验）

H_0: 总体的p^*分位数不大于x^*

[或 H_0: $P(X \leqslant x^*) \geqslant p^*$.]

H_1: 总体的p^*分位数大于x^*

[或 H_1: $P(X \leqslant x^*) < p^*$.]

T_1 的值较小时，表示 H_0 是假的，所以用样本量 n 和特定的概率值 p^* 在表 A3 中得到 t_1，使得

$$P(Y \leqslant t_1) = \alpha \tag{6}$$

对于可以接受的水平 α，其中 Y 服从参数为 n 和 p^* 的二项分布. 如果 T_1 小于等于 t_1，则拒绝 H_0. 如果 T_1 大于 t_1，则接受 H_0. 当 $n>20$ 时，在（1）式中令$q=\alpha$，得 $t_1=x_\alpha$. 138

p-值等于二项随机变量 Y 小于等于观测值 T_1 的概率，当 $n\leqslant 20, p=p^*$ 时，可从表 A3 查出，对于 $n>20$，表 A1 用（4）式，从表 A1 中可得到.

C.（右边检验）

H_0:总体的p^*分位数大于等于x^*

[这等价于 H_0: $P(X<x^*)\leqslant p^*$.]

H_1:总体的p^*分位数小于x^*

[这等价于 H_1: $P(X<x^*)>p^*$.]

由于较大的 T_2 表示零假设是假的，到表 A3 中，用样本量 n 和假设的概率 p^* 作为 p，得到 t_2，使得

$$P(Y>t_2)=\alpha$$

对可接受的显著性水平 α，它等同于

$$P(Y\leqslant t_2)=1-\alpha \tag{7}$$

如果 T_2 大于 t_2，则拒绝 H_0. 如果 T_2 小于等于 t_2，则接受 H_0. 对于 $n>20$，在 1 式中令 $q=1-\alpha$，得到 $t_2=x_{1-\alpha}$.

p-值是二项随机变量 Y 大于等于观测值 T_2 的概率，当 $n\leqslant 20, p=p^*$ 时，它可从表 A3 中查出，对于 $n>20$，用（5）式，它可从表 A1 中查出.

计算机辅助 *Minitab* 在 Median Test 的名义下，可以检验当 $p=1/2$ 的零假设. ——◀

例 3.2.1

大学新生入学后要参加一个特殊的高中学业考试，多年以来成绩的上四分位数是 193. 某个高中有 15 名毕业生上了大学，他们参加了考试，得分如下：

189	233	195	160	212
176	231	185	199	213
202	193	174	166	248

139

认为这 15 个学生是这所高中上大学的所有学生的一个随机样本，比较这所高中毕业的学生和其他大学新生的一个方法就是检验假设：上面所给出的分数来自一个上四分位数是 193 的总体，即

H_0: 上四分位数是193

相应的备择假设是

H_1: 上四分位数不是193

此处，我们讨论这所大学里以前、现在和将来来自这所高中的学生的分数的上四分位数.

用双边分位数检验. 水平大约为 0.05 的临界域可以通过表 A3 查到，此时 $n=15$ 和 $p=0.75$. 可以看出，对于二项随机变量 Y

$$P(Y\leqslant 7)=0.0173 \tag{8}$$

和

$$P(Y \leqslant 14) = 0.9866 = 1 - 0.0134 \tag{9}$$

水平为 α 的临界域

$$\alpha = 0.0173 + 0.0134 = 0.0307 \tag{10}$$

对应于 T_1 小于等于 $t_1 = 7$ 和 T_2 大于等于 $t_2 = 14$.

在这个例子中 $T_1 = 7$，观测值小于等于 193，由于一个观测值严格等于 193，所以 $T_2 = 6$. 因此，因为 T_1 太小，所以拒绝 H_0. 来自那所高中的学生的上四分位数不是 193，p-值是 $2 \cdot P(Y \leqslant 7) = 2(0.0173) = 0.0346$. ■

下面举例说明单边分位数检验和大样本近似.

例 3.2.2

记录了 112 次黄石公园的老忠实间歇泉喷发的间隔时间，要检测间歇时间的中位数是否小于等于 60 分钟（零假设），或中位数是否大于 60 分钟（备择假设）. 如果中位数区间是 60，则 60 是 $x_{0.5}$ 或是中位数. 如果时间间隔的中位数区间小于 60，则 60 是某个 $p \geqslant 0.5$ 的 p 分位数. 所以 H_0 是 $P(X \leqslant 60) \geqslant 0.50$, H_1 是 $P(X \leqslant 60) < 0.50$，其中 X 是喷发的间隔时间. 假设各时间间隔是独立同分布的，可以用左边分位数检验. 检验统计量 T_1 等于间隔时间小于等于 60 的次数，0.05 的临界域，对应于 T_1 小于等于

140

$$\begin{aligned} t_1 &= np^* + z_{0.05}\sqrt{np^*(1-p^*)} \\ &= (112)(0.50) - (1.645)\sqrt{(112)(0.50)(0.50)} = 47.3 \end{aligned} \tag{11}$$

在记录的 112 次时间间隔中，8 个是小于等于 60 分钟的，所以 $T_1 = 8$，H_0 很容易被拒绝，倾向于备择假设"间隔时间的中位数大于等于 60 分钟". 用（4）式得出 p-值.

$$P(Y \leqslant 8) \cong P\left[Z \leqslant \frac{8 - (112)(0.50) + 0.5}{\sqrt{(112)(0.50)(0.50)}}\right] = P(Z \leqslant -8.977) << 0.0001 \tag{12}$$

读为"远小于 0.0001". ■

□**理论**　首先，我们解释为什么 A, B, C 中括号内的假设等价于不在括号内的假设，也许这很容易从一个任意分布函数的图像中看出，如图 3-1.

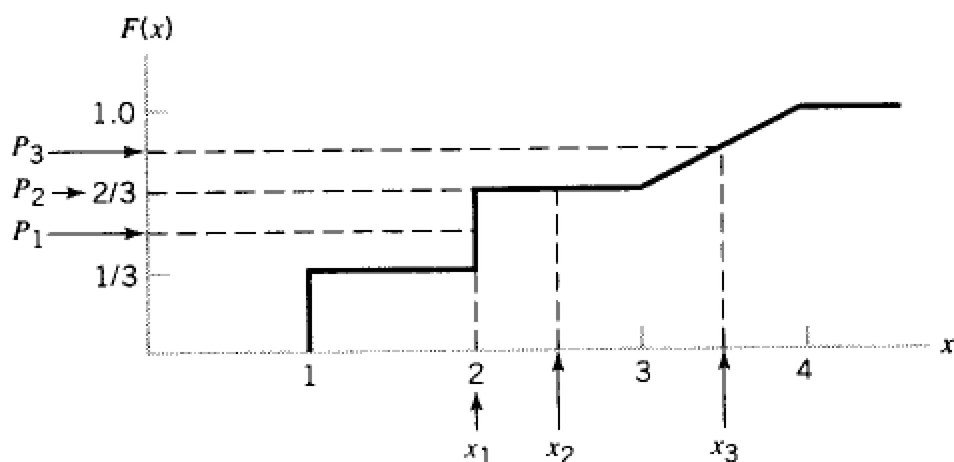


图 3-1　分布函数

分布函数在x^*点有3种状态：它可能像在x_1点一样垂直上升；可能像在x_2点一样为水平线段；或像在x_3点一样逐渐上升. 第2个假设（B假设）的零假设H_0中提出总体的p^*分位数（x_{p^*}）不大于x^*，或者$x_{p^*}\leqslant x^*$. 因为x^*的每一个值都可以看作是某种类型的分位数，我们可以说x^*是对应某个p的p分位数，如p_0.（我们直接从分位数的定义出发，而暂时忽略只选择水平线段的中点作为分位数的习惯.）因为分布函数的图像随着x的增加不会降低，所以$x_{p^*}\leqslant x^*$意味着$p^*\leqslant p_0$，这可以从我们假设x^*是图3-1中由x_1,x_2,x_3代表的三种情况中的任一种看出来. 由任何落在x^*左边的x_{p^*}，可以得出x_{p^*}点的纵坐标p^*不大于x^*点的纵坐标p_0. 由分位数的定义和定义1.4.1，得 141

$$P(X>x^*)\leqslant 1-p_0 \tag{13}$$

和

$$p_0\leqslant 1-P(X>x^*)=P(X\leqslant x^*) \tag{14}$$

等价. 由于$p^*\leqslant p_0$，说明

$$p^*\leqslant P(X\leqslant x^*) \tag{15}$$

这等价于B部分的原假设H_0的形式. H_0的对立假设是H_1，与（15）式对立的是

$$p^*>P(X\leqslant x^*) \tag{16}$$

这正如备择假设所陈述的一样. 同理，可以用来证明其他假设情况的等价形式.

简单地说，图3-1可以形象地说明$x_{p_0}\leqslant x_{p^*}$（C中的H_0）隐含着$p_0\leqslant p^*$. 如果$x^*=x_{p_0}$，那么由定义1.4.1知

$$P(X<x^*)\leqslant p_0\leqslant p^* \tag{17}$$

为真，也是H_0的等价形式.

二项检验可以直接用来检验括号中的零假设. C中的H_0检验可以通过定义二项检验的"第一类"为小于x^*的观测来检验. B中的H_0通过将"第一类"看作小于等于x^*的观测来检验. 结合B和C中的检验就得到A中的双边检验. 二项检验中独立性和概率p为常数的假设是能够成立的，因为X_i是独立同分布的（分别地）. □

在前面章节中我们给出了如何寻找概率p的置信区间，同样的方法可以用来寻找分布函数在某点x_0的值$F(x_0)$的置信区间，即给定x_0，我们能用前面的方法找到未知概率$F(x_0)$的一个"竖直的"置信区间（见图3-1）. 假设给定概率（不妨为p^*），要求未知分位数x_{p^*}"水平态"的置信区间. 如果我们给出一个特定的分位数，如中位数，上四分位数，或任何p^*分位数，其中p^*是个特定的常数，且$0<p^*<1$，则可以找到这种第二类型分位数的置信区间，并有如下所述形式：

$$P(X^{(r)}\leqslant x_{p^*}\leqslant X^{(s)})=1-\alpha \tag{18}$$ 142

其中，$1-\alpha$是已知的置信系数，$X^{(r)},X^{(s)}$是已知第r和第s个次序统计量（见定义2.1.4）. r和s的值可以在只知道样本量和置信系数的情况下，由下一步抽取样本的方式提前决定. 样本$X_1,X_2,\cdots,X_n$只需是随机的，而对X_i的分布没有任何限制. 所

以这个统计方法可以用于任何总体的随机样本.

▶分位数的置信区间

数据 数据由独立同分布的随机变量 $X_1,X_2,\cdots,X_n$ 的观测组成，$X^{(1)}\leqslant X^{(2)}\leqslant\cdots\leqslant X^{(r)}\leqslant\cdots\leqslant X^{(s)}\leqslant\cdots\leqslant X^{(n)}$ 为次序统计量，$1\leqslant r<s\leqslant n$. 希望找到 p^*（未知的）分位数，p^* 是 0 到 1 之间的某个指定数.

假定条件

1. $X_1,X_2,\cdots,X_n$ 是随机样本.
2. X_i 的度量尺度至少是次序的.

方法 A（小样本） 对于 $n\leqslant 20$ 可以用表 A3 来寻找 r 和 s. 在表 A3 中，令 $p=p^*$ 和样本量 n，沿着 $p=p^*$ 的列向下找，直到有近似等于 $\alpha/2$ 的值，此时，$1-\alpha$ 近似于给定的置信系数. 称这个近似值为 α_1，相应的 y 值（远离 α_1 左边）是 $r-1$，加 1 得到 r. 然后继续沿着这列向下找，直到有近似等于 $1-(\alpha/2)$ 的值，称为 $1-\alpha_2$. 对应 $1-\alpha_2$ 的 y 值，记作 $s-1$，加 1 得到 s. 这样我们得到了 α_1,α_2,r,s. 准确置信系数是 $1-\alpha_1-\alpha_2$，区间估计量是 $X^{(r)},X^{(s)}$ 之间的区间，而 $X^{(r)},X^{(s)}$ 可以从数据中得到. 那么，

$$P(X^{(r)}\leqslant x_{p^*}\leqslant X^{(s)})\geqslant 1-\alpha_1-\alpha_2 \tag{19}$$

给出了置信区间. 如果假设未知的分布函数是连续的，那么，

$$P(X^{(r)}\leqslant x_{p^*}\leqslant X^{(s)})=1-\alpha_1-\alpha_2 \tag{20}$$

和（18）式所述的一样.

143 **方法 B（大样本近似）** 对于 n 大于 20，可以用基于中心极限定理的逼近（见（1）式），计算

$$r^*=np^*+z_{\alpha/2}\sqrt{np^*(1-p^*)} \tag{21}$$

和

$$s^*=np^*+z_{1-\alpha/2}\sqrt{np^*(1-p^*)} \tag{22}$$

其中，分位数 z_q 从表 A1 找出，$1-\alpha$ 是给定的置信系数. 一般地，r^*,s^* 不是整数. 令 r 和 s 是 r^*,s^* 向上取整的整数. 则（19）式给出了近似置信区间，如果未知分布函数是连续的，则近似置信区间由（20）式给出.

像所叙述的一样，单边的置信区间可以通过只找 r 或 s 得到. 单边置信区间有形式

$$P(X^{(r)}\leqslant x_{p^*})=1-\alpha_1 \tag{23}$$

和

$$P(x_{p^*}\leqslant X^{(s)})=1-\alpha_2 \tag{24}$$

如果分布函数是连续的，则为

$$P(X^{(r)} \leqslant x_{p^*}) \geqslant 1-\alpha_1 \tag{25}$$

和

$$P(x_{p^*} \leqslant X^{(s)}) \geqslant 1-\alpha_2 \tag{26}$$

计算机辅助　*Minitab*（在 Median Test 下）和 *StatXact*（在 Sign Test 下）可以得出中位数的置信区间. ◀

例 3.2.3

从一批待测晶体管中随机抽出 16 个进行检测，记录每个晶体管的寿命. 希望得到上四分位数的置信系数近似 90% 的置信区间. 查表 A3，$n=16, p=0.75$. 沿 $p=0.75$ 列向下找，选择概率 0.0271，因为它接近 0.05，对应 $\alpha_1=0.0271$ 的 y 值为 $y=8$；所以 r 等于 9. 最接近 0.95 的概率是 $0.9365=1-\alpha_2$，它对应的 y 值为 14，所以 s 等于 15. 置信区间为 144

$$P(X^{(9)} \leqslant x_{0.75} \leqslant X^{(15)}) = 0.9094 \tag{27}$$

(可以假定寿命是连续随机变量，所以可以用（20）式.

以递增的顺序列出检验的结果，如下：

$X^{(1)}=46.9$	$X^{(5)}=56.8$	$X^{(9)}=63.3$	$X^{(13)}=67.1$
$X^{(2)}=47.2$	$X^{(6)}=59.2$	$X^{(10)}=63.4$	$X^{(14)}=67.7$
$X^{(3)}=49.1$	$X^{(7)}=59.9$	$X^{(11)}=63.7$	$X^{(15)}=73.3$
$X^{(4)}=56.5$	$X^{(8)}=63.2$	$X^{(12)}=64.1$	$X^{(16)}=78.5$

由于 $X^{(9)}=63.3, X^{(15)}=73.3$，我们可以说"从 63.3 小时到 73.3 小时的区间是上四分位数的 90.94% 置信区间."

由（21）和（22）式，用大样本近似，得到

$$\begin{aligned} r^* &= (16)(0.75)+(-1.645)\sqrt{(16)(0.75)(0.25)} \\ &= 12-2.86=9.14 \end{aligned} \tag{28}$$

和

$$s^* = 12+2.86=14.86 \tag{29}$$

从而，$r=10, s=15$，90% 置信区间是（63.4, 73.3），比精确方法得到的区间略小. ■

□**理论**　首先考虑分布函数是连续的这种较为简单的情况. 如果 x_{p^*} 是 p^* 分位数，则有如下严格关系

$$P(X \geqslant x_{p^*}) = P(X > x_{p^*}) = 1-p^* \tag{30}$$

其中，X 的分布函数和随机样本的分布函数一样.

次序统计量 $X^{(1)}$，假设大于某个确定的常数，只要样本中最小的数都大于这个常数，所以只要样本中 n 个值都大于这个常数，$X^{(1)}$ 就大于这个常数. 选择 x_{p^*} 作为这个常数，可以得到

$$\begin{aligned} P(x_{p^*} < X^{(1)}) &= P(\text{所有的样本值都大于 } x_{p^*}) \\ &= P(x_{p^*} < X_1, x_{p^*} < X_2, \ldots, x_{p^*} < X_p) \\ &= P(x_{p^*} < X_1) \cdot P(x_{p^*} < X_2) \cdot \cdots \cdot P(x_{p^*} < X_p) \\ &= (1 - p^*)^n \end{aligned} \tag{31}$$

145 因为 X_i 是独立的，它们都有同样的 p^* 分位数 x_{p^*}.

如果 x_{p^*} 小于 $X^{(2)}$，那么在 $X^{(1)} \leqslant x_{p^*} < X^{(2)}$ 中，恰有 $n-1$ 个观测值大于 x_{p^*}，或者在 $x_{p^*} < X^{(1)} < X^{(2)}$ 中，有 n 个观测值大于 x_{p^*}. 所以

$$\begin{aligned} P(x_{p^*} < X^{(2)}) &= P(x_{p^*} < X^{(1)}) + P(X^{(1)} \leqslant x_{p^*} < X^{(2)}) \\ &= P(X_i \text{中至少有 } n-1 \text{ 大于 } x_{p^*}) \\ &= P(X_i \text{中至多有1个} \leqslant x_{p^*}) \end{aligned} \tag{32}$$

现在，(32) 式中的概率由二项分布函数给出，因为每个 X_i 都有小于等于 x_{p^*} 的概率 p^*，且 X_i 是互相独立的. 所以由 (32) 式可以得到

$$P(x_{p^*} < X^{(2)}) = \sum_{i=0}^{1} \binom{n}{i} (p^*)^i (1-p^*)^{n-i} \tag{33}$$

在 (1.3.8) 式中的二项分布函数之下，之前的讨论可以作如下推广，

$$\begin{aligned} P(x_{p^*} < X^{(r)}) &= P(X_i \text{中至少有} n-r+1 \text{大于} x_{p^*}) \\ &= P(X_i \text{中至多有} r-1 \text{个} \leqslant x_{p^*}) \\ &= \sum_{i=0}^{r-1} \binom{n}{i} (p^*)^i (1-p^*)^{n-i} \end{aligned} \tag{34}$$

置信系数由下式得出

$$\begin{aligned} 1-\alpha &\cong P(X^{(r)} \leqslant x_{p^*} \leqslant X^{(s)}) \\ &= P(x_{p^*} \leqslant X^{(s)}) - P(x_{p^*} < X^{(r)}) \end{aligned} \tag{35}$$

从而，由 (34) 式和表 A3 可以得到 r, s，使得

$$1-\alpha_2 = P(x_{p^*} \leqslant X^{(s)}) \cong 1 - \frac{\alpha}{2} \tag{36}$$

和

$$\alpha_1 = P(x_{p^*} < X^{(r)}) \cong \frac{\alpha}{2} \tag{37}$$

则置信系数是 $1-\alpha_1-\alpha_2 \cong 1-\alpha$. 注意，因为假设分布函数是连续的，我们有

$$P(x_{p^*} \leqslant X^{(s)}) = P(x_{p^*} < X^{(s)}) \tag{38}$$

146 因此，可以用表 A3 得到 s.

如果 X 的分布函数和 X_i 的分布函数不是连续的，(30) 式不成立. 由定义 1.5.1，我们有

$$P(X > x_{p^*}) \leqslant 1 - p^* \tag{39}$$

和

$$P(X \geqslant x_{p^*}) \geqslant 1 - p^* \tag{40}$$

首先，我们考虑（39）式怎样影响（34）式，进而影响（37）式求 r 的方法. 因为（39）式成立，每个观测值大于 x_{p^*} 的概率小于当 X 是连续时的值，所以，每个次序统计量大于 x_{p^*} 的倾向，小于 X 为连续时的情形. 即概率 $P(x_{p^*}<X^{(r)})$ 小于连续时（34）式给出的值. 所以，一般情况下，下式成立

$$P(x_{p^*}<X^{(r)})\leqslant\sum_{i=0}^{r-1}\binom{n}{i}(p^*)^i(1-p^*)^{n-i} \tag{41}$$

如果用上面介绍的方法从表 A3 中找 r，那么

$$P(x_{p^*}<X^{(r)})\leqslant\alpha_1 \tag{42}$$

现在，我们来考虑（40）式怎样影响通过选择 s 的值得到概率 $1-\alpha_2$ 的. 因为（40）式成立，每个观测值大于等于 x_{p^*} 的概率大于连续时的概率，所以观测值大于等于 x_{p^*} 的个数比连续时多，$X^{(s)}\geqslant x_{p^*}$ 的概率大于连续时的情况. 因此，（34）式可以改为适应一般情况的式子.

$$P(x_{p^*}\leqslant X^{(s)})\geqslant\sum_{i=0}^{s-1}\binom{n}{i}(p^*)^i(1-p^*)^{n-i} \tag{43}$$

所以，如果用先前的方式在表 A3 中找 s，我们有

$$P(x_{p^*}\leqslant X^{(s)})\geqslant 1-\alpha_2 \tag{44}$$

对于任何分布都成立的（42）和（44）式，可以按如下方式使用

$$\begin{aligned}P(X^{(r)}\leqslant x_{p^*}\leqslant X^{(s)}) &= P(x_{p^*}\leqslant X^{(s)})-P(x_{p^*}<X^{(r)})\\ &\geqslant P(x_{p^*}\leqslant X^{(s)})-\alpha_1\\ &\geqslant 1-\alpha_2-\alpha_1\end{aligned} \tag{45}$$ 147

所以，这种方法对于离散随机变量或有结点的有序数据是保守的. 因此，求分位数的置信区间的方法，对有二项分布函数的精确表可用的情形也是可行的.

用大样本方法求 r 和 s 是基于用标准正态分布近似二项分布的想法，虽然关于怎样由 r^*,s^* 求得整数 r,s 的方法还有不同的争论，但是，此处给出的直接向上取整的方法是个很接近的近似.

这种分位数检验可以用于处理次序数据，因此比其他参数检验方法更适用. 如果数据是以区间为单位的，且服从正态分布，均值等于中位数，中位数的分位数检验可以比作一样本的 t 检验，此时渐近相对效率（A. R. E.）只为 $2/\pi=0.637$. 对于均匀分布，它是轻尾的，它的 A. R. E. 只有 $1/3=0.333$. 但是，对称的重尾分布，如我们所知的双指数分布，分位数检验相对于 t 检验的 A. R. E. 就跳到了 2.0，说明对于非正态的重尾分布，分位数检验比参数检验更有效. □

Barlow 和 Gupta(1966)讨论了分位数的单边置信区间在寿命检验中的应用，Van der Parren(1970)给出了中位数分布自由置信限的表和分位数的表（1973），Krewski(1976)和 Reiss 和 Rüschendorf(1976)讨论了分位数之间的区间的置信区间.

习题

1. 一个 10 年级学生体重的随机样本有如下 20 个观测值

142	134	98	119	131	103	154	122	93	137
86	119	161	144	158	165	81	117	128	103

检验假设：体重的中位数是 103.

2. 在题 1 中检验假设：上四分位数至少是 150.
3. 在题 1 中检验假设：30% 分位数不大于 100.
4. 在题 1 中求中位数的近似 90% 的置信区间. 准确的置信系数是多少？比较用准确方法得
148 到的结果和用近似方法得到的结果.
5. 要设计某种汽车的车内高度以适应大部分司机，除了那些占 5% 的超高司机之外，以前的研究表明 95% 分位点是 70.3 英寸. 为了验证以前的研究是否仍然有效，选择 100 个随机样本，发现样本中最高的 12 个人有如下高度：

72.6	70.0	71.3	70.5	70.8	76.0
70.1	72.5	71.1	70.6	71.9	72.8

用 70.3 作为 95% 分位点合理吗？

6. 在习题 5 中，样本的 95% 分位点的 95% 的置信区间是什么？
7. 警官回忆说，当年完成超越障碍训练需要 42 分钟，他怀疑现在入伍的新兵是否能达到当时新兵的标准，所以他记录了他们完成超越障碍训练的时间. 他发现在 38 名新兵中只有 10 个在 41 分钟内完成了训练. 用分位数检验假设：上四分位数是 42 分钟，相应的备择假设是单边的.
8. 检验 10cm 的钢板能被子弹穿进多深. 50 发子弹射向钢板，测量它们进入钢板的深度，7 发子弹穿透钢板，所以它们的射进深度为记为 10 +，所有 50 个深度由小到大如下所示：

5.37, 5.39, 5.42, 5.51, 5.63, 5.74, 5.82, 5.83, 5.94, 5.98, 6.07, 6.07, 6.13, 6.20, 6.21, 6.23, 6.25, 6.26, 6.26, 6.28, 6.29, 6.31, 6.35, 6.41, 6.57, 6.67, 6.81, 7.03, 7.40, 7.44, 7.82, 8.03, 8.11, 8.44, 8.51, 8.72, 8.83, 9.04, 9.33, 9.51, 9.61, 9.68, 9.82, 10+, 10+, 10+, 10+, 10+, 10+, 10+

求射进深度的中位数的 95% 置信区间.

思考题

一种求中位数的 $1-\alpha$ 置信区间的参数方法是假设总体服从正态分布，用

$$\overline{X} + t_{\alpha/2}S/\sqrt{n-1} < x_{0.5} < \overline{X} + t_{1-\alpha/2}S/\sqrt{n-1}$$

其中 $\bar{X}$ 是样本均值，S 是样本标准差（定义 2.2.3），n 是样本量，t_p 是表 A21 中的 p 分位数，自由度为 $n-1$. 计算习题 1 中数据的置信区间，将它与习题 4 的非参数 $\alpha=0.10$ 置信区间比较一下，哪个置信区间更容易证明？哪个置信区间“更好”（在更短的意
149 义下）？

3.3 容 忍 限

3.1 和 3.2 节的置信区间给出了总体未知参数的估计，如未知概率 p 或未知分位数 x_p，以及未知参数在某区间内的 $1-\alpha$ 概率（置信系数）. 容忍限不同于置信区间，因为容忍限给出总体比例至少为 q 的所在的区间，使得此区间“确实”含有总体比例 q 的概率大于等于 $1-\alpha$. 典型的应用就是我们抽取容量为 n 的随机样本 $X_1,X_2,\cdots,X_n$，要想知道 n 需要多大，才能使我们有 95% 的把握说总体至少有 90% 落在 $X^{(1)}$，$X^{(n)}$之间，其中 $X^{(1)},X^{(n)}$ 为样本的最小和最大值. 我们可以进一步推广或考虑问题，“n 至少要多大，才能使总体至少有 q 的比例落在 $X^{(r)},X^{(n+1-m)}$之间的概率大于等于 $1-\alpha$?” 其中，数 q,r,m，α 是事先已知（或选取）的，只需要确定 n.

另一个典型的情况是，当有 n 个随机样本时，希望选取上下限使得有 95% 置信度（或 $1-\alpha$）说，我们所选择的置信限包含总体的比例至少为 q. 如果我们选取样本的两个极值 $X^{(1)},X^{(n)}$ 为上下限，那么总体的比例 q 是多少？或者我们还是选样本的第二极值 $X^{(2)},X^{(n-1)}$ 为上下限？在 95% 置信水平下，总体有多大的比例在这些限内？在这个问题中，q 是未知量，并且当我们知道或设置 α,n,r,m 后可以得到它.

上面的容忍限是双边容忍限. 单边容忍限一般有形式，“总体至少有 q 的比例大于 $X^{(r)}$的概率为 $1-\alpha$,” 或 “总体至少有 q 的比例小于 $X^{(n+1-m)}$ 的概率为$1-\alpha$.” 单边容忍限与分位数的单边置信区间是一样的，本节将在下面介绍.

此处所说的总体是无穷的或抽样是有放回的，以使得 X_i 是独立的，对于有限样本，其抽样是不放回的，且样本容量 n 与总体样本 N 相比很小，这些方法是相当准确的. 对于有限总体来说，更精确的方法可参见 Wilks(1962).

▶容忍限

数据 数据包含来自一个很大总体的随机样本 $X_1,X_2,\cdots,X_n$，选择一个置信系数 $1-\alpha$ 和一对正整数 r,m. 我们要在选定理想的总体比例 q 之后确定所需样本容量（见方法 A)，或要对于给定的样本容量 n，再确定总体比例 q（见方法 B). 给出一个陈述， 150
“从 $X^{(r)}$ 到 $X^{(n+1-m)}$ 的随机区间里包含总体比例 q 或更多样本的概率是 $1-\alpha$.” 注意，我们约定 $X^{(0)}=-\infty,X^{(n+1)}=+\infty$，所以单边容忍限可以通过令 r 或 m 为零得到.

假定条件

1. $X_1,X_2,\cdots,X_n$ 是一组随机样本.
2. 度量尺度至少是须序的.

方法 A（求 n） 如果 $r+m$ 等于 1，即如果 r 或 m 等于零，就像在单边容忍限一样，对合适的 α,q 值，可直接从表 A5 中得到 n. 如果 $r+m$ 等于 2，对合适的 α,q 值，可直接从表 A6 中获得 n. 如果表 A5 和 A6 都不行，则用下面的近似

$$n\cong\frac{1}{4}x_{1-\alpha}\frac{1+q}{1-q}+\frac{1}{2}(r+m-1) \tag{1}$$

其中，$x_{1-\alpha}$是自由度为$2(r+m)$的χ^2随机变量的（$1-\alpha$）分位数，由表 A2 得到.

方法 B（求 q）　对于已知的样本量 n，已定的 α，r，m，总体比例 q 的近似值由下式给出

$$q=\frac{4n-2(r+m-1)-x_{1-\alpha}}{4n-2(r+m-1)+x_{1-\alpha}} \tag{2}$$

其中，$x_{1-\alpha}$是自由度为$2(r+m)$的χ^2随机变量的（$1-\alpha$）分位数（由表 A2 得到）.

容忍限　对样本容量为 n，总体至少有 q[或$(100)(q)\%$]的比例落在 $X^{(r)}$ 到 $X^{(n+1-m)}$ 之间的概率至少为 $1-\alpha$，即

$$P(X^{(r)}\leqslant \text{总体至少有}q\text{的比例}\leqslant X^{(n+1-m)})\geqslant 1-\alpha \tag{3}$$

对于单边容忍区域，令 r 或 m 等于零，其中 $X^{(0)}=-\infty, X^{(n+1)}=+\infty$，可用类似于上面的过程获得. ◀

例 3.3.1

151 使用最广泛的双边容忍限是 $r=1, m=1$. 在某流行的豪华轿车中配有电动座位调节器，制造商想了解调节的高度范围，使得在 90% 的概率下，至少有 80% 的潜在买主（总体）能够调节座位到理想的高度. 样本容量 n 是多少时，才能使 $X^{(n)}, X^{(1)}$ 分别是容忍限的上下限？

在表 A6 中令 $q=0.80, 1-\alpha=0.90$，则得到 n 是 18. 由（1）式近似得到

$$\begin{aligned} n &\cong \frac{1}{4}x_{1-\alpha}\frac{1+q}{1-q}+\frac{1}{2}(r+m-1) \\ &= \frac{1}{4}(7.779)\frac{1.80}{0.20}+\frac{1}{2}=18.003 \end{aligned}$$

从潜在买主中抽取 18 个人作为一个样本，测量从一个基准高度起所调节的高度. 样本中最大值是

$$X^{(18)}=7.57\text{ 英寸}$$

最小值是

$$X^{(1)}=1.21\text{ 英寸}$$

所以，至少有 80% 的人需要调节座位垂直高度等于或在 1.21 和 7.57 英寸之间的概率为 0.9. ■

下面是一个单边容忍限的例子.

例 3.3.2

在一些钢筋中，制造商保证每批至少有 90% 的钢筋有一个超过额定数的断裂临界点. 因为制造环境不同，通过找每批随机样本的断裂点，来分批建立所保证的断裂临界点，让所保证的断裂临界点等于样本的最小断裂临界点. 需要多大的样本量，制造商才能有 95% 的把握说他们的保证是正确的？

在表 A5 中，令 $q=0.90, 1-\alpha=0.95$，得到 n 是 29. 每批随机抽取 29 个样本，样本中最小的断裂临界点就是保证的断裂临界点，即这一批中至少 90% 的钢筋会以 95% 的概率保持完好无损. ■

例 3.3.3

一批桶是用来安全存放有放射性的垃圾，每一桶都标有所含放射性垃圾的量，并对此进行周期性的检查，从中随机地抽取几桶并从外部扫描，来估计桶中放射性垃圾的含量，估计值与桶上标签的值相比较得到差异 X. 过了 3 个月的周期后，用这种方法检查了 122 桶，结果得到随机变量 $X_1,\cdots,X_{122}$，其中 X_i 是桶上标的数量和扫描估计量的差值. 152

选取上下限 $X^{(2)},X^{(121)}$（即 $r=2,m=2$）以及置信水平 95%. 由（2）式得到落在这个区间内总体的比例，用自由度为 $2(r+m)=2(2+2)=8$ 的 χ^2 分布的 0.95 分位数，从表 A2 中查得为 15.51.

$$q=\frac{4n-2(r+m-1)-x_{1-\alpha}}{4n-2(r+m-1)+x_{1-\alpha}}=\frac{488-6-15.51}{488-6+15.51}=0.938$$

我们可以有 95% 的把握说，122 桶中至少有 93.8% 的桶的差异在第二小观测值和第二大观测值之间. ■

□**理论** 仔细检查单边容忍限所做的表述，可以发现它与单边分位数置信区间的相似性，即单边容忍限指的是：

$$P(\text{总体至少有}q\text{的比例}\leqslant X^{(n+1-m)})\geqslant 1-\alpha \tag{4}$$

但是，“总体至少有 q 的比例$\leqslant X^{(n+1-m)}$” 与 “总体 q 分位数$\leqslant X^{(n+1-m)}$” 是一样的；这两种表述只是在表述想法上有所不同. 所以，我们有

$$\begin{aligned}&P(\text{总体至少有}q\text{的比例}\leqslant X^{(n+1-m)})\\&=P(\text{总体 }q\text{ 分位数}\leqslant X^{(n+1-m)})=P(x_q\leqslant X^{(n+1-m)})\end{aligned} \tag{5}$$

（5）式中的概率在（3.2.43）式中已给出

$$P(x_q\leqslant X^{(n+1-m)})\geqslant\sum_{i=0}^{n-m}\binom{n}{i}q^i(1-q)^{n-i} \tag{6}$$

检查（6）式的右边发现，它用来求使（6）式右端超过 $1-\alpha$ 的最小 n，这个可以通过在表 A3 中令 $y=n-m$，参数 p 等于 q，然后寻找最小值 n，使得其值大于等于 $1-\alpha$. 因为当 y 的值随 n 的变化而变化时，为方便起见，将（6）式的右边改写为 153

$$\sum_{i=0}^{n-m}\binom{n}{i}q^i(1-q)^{n-i}=1-\sum_{i=n-m+1}^{n}\binom{n}{i}q^i(1-q)^{n-i} \tag{7}$$

是可以的，因为所有二项概率的和等于 1. 变化（7）式右端的指标 $j=n-i$，可得

$$\sum_{i=0}^{n-m}\binom{n}{i}q^i(1-q)^{n-i}=1-\sum_{j=0}^{m-1}\binom{n}{j}(1-q)^jq^{n-j} \tag{8}$$

（8）式从事实“$n-m$ 或更少次数成功的概率等价于 m 或更多次失败的概率，它等于 1 减 $m-1$ 次或更少次失败的概率”中立即得出. 结合（8）和（6）式，我们可以找到最小的 n，通过使它满足

$$\sum_{j=0}^{m-1}\binom{n}{j}(1-q)^j q^{n-j}\leqslant\alpha \tag{9}$$

这等价于不等式

$$1-\sum_{j=0}^{m-1}\binom{n}{j}(1-q)^j q^{n-j}\geqslant 1-\alpha \tag{10}$$

则在表 A3 中，令 $y=m-1, p=1-q$，直到找到有值小于等于 α. 这个对应的 n 值就是所选取的样本容量.

另一个单边容忍限是

$$P(\text{总体中至少有}q\text{ 的比例}\geqslant X^{(r)})\geqslant 1-\alpha \tag{11}$$

等价于下式

$$P(X^{(r)}\leqslant x_{1-q})\geqslant 1-\alpha \tag{12}$$

因为总体至少有 $1-q$ 的比例大于等于 x_{1-q}. （12）式就成为

154 $$\alpha\geqslant 1-P(X^{(r)}\leqslant x_{1-q})=P(x_{1-q}<X^{(r)}) \tag{13}$$

从（3.2.41）式可以看出（13）式的解是满足下式的最小 n

$$\alpha\geqslant\sum_{i=0}^{r-1}\binom{n}{i}(1-q)^i q^{n-i} \tag{14}$$

正如（9）式那样.

事实上，通过微积分（见 Noether, 1967a）可以证明，对双边容忍限和两种单边容忍限，样本容量 n 依据不等式

$$\alpha\geqslant\sum_{i=0}^{r+m-1}\binom{n}{i}(1-q)^i q^{n-i} \tag{15}$$

求解.（15）式仅依赖于 $r+m$ 的和，而不依赖于我们是否想选所有的值都在 $X^{(r+m)}$ 右边的区间，或所有的值都在 $X^{(n+1-r-m)}$ 左边的区间，或所有值都在 $X^{(r)}$ 和 $X^{(n+1-m)}$ 之间的区间，或由任何两个次序不同于 $n+1-m-r$ 的次序统计量所组成的区间，这有些令人惊奇.

一般用表 A3 解(15)式是无效的. 所以表 A5 和 A6 给出了最常用的 $r+m=1$ 和 $r+m=2$ 时的值. Scheffé 和 Tukey(1944)不加证明地给出了(1)式的逼近，(2)式可由对 q 解(1)式而获得. Murphy(1984)和 Birnbaum 和 Zuckerman(1949)给出用图形来帮助寻找 n. □

容忍限也可以用于两个样本（Danziger 和 Davis, 1964），用于一个删失的样本（Bohrer, 1968），或用于判定一个样本来自两个可能的多元总体的哪一个（Quesenberry 和 Gessaman, 1968）. Hanson 和 Owen(1963)检验了容忍限在离散随机变量上的使用. Bowden(1968)讨论了容忍限在回归问题中的应用. Mack(1969)以及 Goodman 和 Madansky(1962)发表了其有关容忍限的文章.

习题

1. 以 90% 的概率认为至少有 95% 的总体落在样本极差中，则需要多大的样本量？ 155
   (a) 用精确表格. (b) 用近似方法.
2. 以 95% 的把握说至少有 90% 的总体大于等于 $X^{(1)}$，需要多大的样本量？
   (a) 用精确表格. (b) 用近似方法.
3. 使得至少有 85% 的总体 $\leqslant X^{(n)}$ 的概率为 0.90，则样本量必须是多少？
   (a) 用精确表格. (b) 用近似方法.
4. 使得至少有 99% 的总体 $\geqslant X^{(2)}$ 的概率为 95%，样本量必须是多少？
   (a) 用精确表格. (b) 用近似方法.
5. 至少有 50% 的总体在 $X^{(5)}$ 和 $X^{(n-4)}$ 之内的概率为 0.90，样本量必须是多少？
6. 习题 5 中，如果把概率 0.90 换为 0.95，那么样本容量必须是多少？
7. 健身中心测量了 86 个会员的含脂肪比例.
   (a) 在 95% 的概率下，样本中 86 个会员的含脂肪比例在最小比例和最大比例之间的人数比例最小是多少？在 90% 的概率下，情况如何呢？
   (b) 在 95% 的概率下，样本中 86 个会员的脂肪比例在 $X^{(2)}$ 和 $X^{(85)}$ 之间的人数比例最小是多少？在 90% 的概率下，情况如何呢？
8. 目录邮购公司通过平信调查了它的 146 个顾客，来了解它最近订单的交货周期（从下订单日期到交付日期）.
   (a) 在 95% 的概率下，由样本观测值得到的顾客期望邮递时间在 $X^{(1)}$ 和 $X^{(142)}$ 之间的比例至少是多少？对 90% 概率的情况呢？
   (b) 注意在这种情况下容忍区间的端点是不对称的. 在这个问题中用这些不对称的端点有什么优点？
9. 某工程师记录了收到的一批不锈钢杆的规格，发现至少 90% 的杆长在她随机选的杆长的第 6 长和第 6 短之间，为了得到这个表述的 99% 的可信度，样本量应当是多大？
10. 研制了一个计算机模型来模拟战争中一个作战单位（例如一个通讯中心），其中由计算机模型决定的重要一项，就是保持作战单位满意运转水平的最少人数. 我们希望给作战单位配备足够多的人，使得 90% 的战斗它都能满意地运转.
    (a) 计算机需要运行多少次，才能使得我们有 99.9% 的把握说需要的人数不多于 $X^{(n)}$，运行的观测值的最大数是多少？
    (b) 计算机需要运行多少次，才能使得我们有 99.9% 的把握说需要的人数在 $X^{(1)}$ 和 $X^{(n)}$ 之间？ 156
    (c) 计算机需要运行多少次，才能使得我们有 99.9% 的把握说需要的人数在 $X^{(2)}$ 和 $X^{(n-1)}$ 之间？
    (d) 计算机需要运行多少次，才能使得我们有 99.9% 的把握说需要的人数不多于 $X^{(n-4)}$？

思考题

用表 A3 求解习题 3，求准确的 α 值.

3.4　符号检验

在前面的一节中，我们的主题有些偏离假设检验，现在重新回来讨论最古典的非参数检验，即符号检验. 实际上，符号检验就是参数为 $p^* = 1/2$ 的二项检验. 但是，因为它的使用广泛性和经典性（可追溯到1710年），符号检验值得更特别考虑，还因为 $p^* = 1/2 = 1 - p^*$，使它比二项检验更加简单. 符号检验常用于检验一对变量 (X,Y) 中的一个随机变量是否比另一随机变量大. 同时，我们将在 3.5 节中用它来检验一系列次序度量的趋势，或检验相关性. 对同一模型，在很多可以用符号检验的情况下，也可以用更有效的非参数检验. 但是符号检验通常用起来更简单和方便，求临界域经常不需要特殊的表.

▶符号检验

数据　数据是一组二维随机样本 $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$ 的观测，其中有 n' 对观测. 对配对观测来说，应当有某个自然基础；否则 X 和 Y 是独立的，这时更适合使用第 5 章更有效的 Mann-Whitney 检验.

在每对 (X_i,Y_i) 之间进行比较，如果 $X_i < Y_i$，记为"+"或"正"；如果 $X_i > Y_i$，记为"-"，或"负"；如果 $X_i = Y_i$，记为"0"或"结点". 因此只需要度量是次序的.

假定条件

1. 二维随机变量 $(X_i,Y_i), i=1,2,\cdots,n'$，是相互独立的.

2. 每对之间的度量尺度是须序的. 即每对 (X_i,Y_i) 必定是"正"，"负"或
157 "结点"中的一个.

3. (X_i,Y_i) 是内部相容的，因为如果对于一对 (X_i,Y_i)，有 $P(+) > P(-)$，那么所有对都有 $P(+) > P(-)$. $P(+) < P(-)$ 和 $P(+) = P(-)$ 的情况也一样.

检验统计量　令检验统计量 T 为"正"的对数；即 T 等于 X_i 小于 Y_i 的 (X_i,Y_i) 对数.

$$T = \text{"+" 的总个数}$$

零分布　T 的零分布服从参数为 $p = 1/2, n =$ 非结点对数的二项分布. 即不考虑所有有结点的对 (X,Y)，其中 $(X=Y)$，且令

$$n = \text{"+" 的总个数和 "-" 的总个数}$$

假设

A. （双边检验）

$$H_0: P(+) = P(-)$$
$$H_1: P(+) \neq P(-)$$

对于 $n \leqslant 20$，用 $p = 1/2$，查表 A3，在表中选择一个大约等于 $\alpha/2$ 的值称为 α_1，对应

于 α_1 的 y 值称为 t. $2\alpha_1$ 水平的临界域对应着 T 值小于等于 t，或大于等于 $n-t$. 如果 $T\leqslant t$ 或 $T\geqslant n-t$，以 $2\alpha_1$ 的显著性水平拒绝 H_0，否则接受 H_0.

对于 n 大于 20，用表 A3 最后的正态逼近，得到

$$t=\frac{1}{2}(n+z_{\alpha/2}\sqrt{n}) \tag{1}$$

其中，$z_{\alpha/2}$ 由表 A1 查得. 如果 $\alpha=0.05$，$z_{\alpha/2}=(-1.96)$，(1)式近似变成

$$t=\frac{n}{2}-\sqrt{n} \tag{2}$$

这很容易记住.

对于 Y 小于等于观测值 T 的概率及 Y 大于等于观测值 T 的概率，p-值是 2 倍于这两个概率中的较小者，对 $n\leqslant 20$，它可以从表 A3 中用 $p=1/2$ 得到，或对 $n>20$， 158
从表 A1 中用

$$P(Y\leqslant t_{\text{obs}})=P\left(Z\leqslant\frac{2\cdot t_{\text{obs}}-n+1}{\sqrt{n}}\right) \tag{3}$$

和

$$P(Y\geqslant t_{\text{obs}})=1-P\left(Z\leqslant\frac{2\cdot t_{\text{obs}}-n-1}{\sqrt{n}}\right) \tag{4}$$

获得，其中，因子 1.0 作为“连续性修正”来改进二项分布的正态分布逼近效果.

B.（左边检验）

$$H_0: P(+)\geqslant P(-)$$
$$H_1: P(+)<P(-)$$

T 值较小说明更可能是“负”而不是“正”，符合 H_1. 在表 A3 中用 $p=1/2$ 和 n，查表得到的近似 α 值，比如是 α_1，则对应于 α_1 的 y 值就是 t. 当 n 大于 20，t 可以通过下面的近似式得到

$$t=\frac{1}{2}(n+z_{\alpha}\sqrt{n}) \tag{5}$$

其中，z_α 从表 A1 中获得.

水平 α_1（或 α）的临界域对应着 T 值小于等于 t. 如果 $T\leqslant t$，则以显著性水平 α_1（或 $n>20$ 时的 α）拒绝 H_0，否则接受 H_0.

p-值为 Y 小于等于观测值 T 的概率，对于 $n\leqslant 20$，用 $p=0.5$ 可以从表 A3 中得到它，或对于 $n>20$，用（3）式，在表 A1 中查得.

C.（右边检验）

$$H_0: P(+)\leqslant P(-)$$
$$H_1: P(+)>P(-)$$

较大的 T 值说明更可能是“正”而不是“负”，正如 H_1 的表述. 所以临界域对应着 T 值大于等于 $n-t$，其中 t 可以通过在表 A3 中用 $p=1/2$ 和 n 去查找近似等于 α 的值（如 159
左边检验的情形）而得到，对应 y 的值就是 t. 对于 n 大于 20，t 可以通过（5）式近似得到. 所以，如果 T 大于等于 $n-t$，则以显著性水平 α 拒绝 H_0.

p-值为 Y 大于等于观测值 T 的概率，对 $n \leqslant 20$，用 $p=0.5$，它可以从表 A3 得到，或对于 $n>20$，用（4）式，在表 A1 中获得.

应当注意，当检验这些假设时，符号检验是无偏和相合的. 符号检验也可以用来检验下面的其他假设，在这种情况下，除非对于（X_i, Y_i）的分布有限制，否则既不是无偏的也不是相合的.

A.（双边检验）

零假设为"X_i 和 Y_i 有相同的位置参数"，所以

$$H_0: E(X_i)=E(Y_i) \quad 对于所有的 i,$$

对备择假设：

$$H_1: E(X_i) \neq E(Y_i) \quad 对于所有的 i.$$

看 X_i 和 Y_i 是否有不同的均值. 这样的检验也可类似地用于对中位数的检验.

$$H_0: 对于所有的 i, X_i 和 Y_i 的中位数相等$$
$$H_1: 对于所有的 i, X_i 和 Y_i 的中位数不相等$$

B.（左边检验）

考虑前面所述的 B 方法，其零假设表明 X_i 的取值倾向可能比 Y_i 来得小；所以这个单边符号检验可以用来检验

$$H_0: E(X_i) \leqslant E(Y_i) \quad 对于所有的 i$$

对备择假设：

$$H_1: E(X_i) > E(Y_i) \quad 对于所有的 i$$

对于中位数检验也有类似的假设表述.

C.（右边检验）

160 由于 H_0 表明 X_i 很可能大于 Y_i 而不是小于 Y_i，因此考虑 X_i 的取值趋向大于 Y_i 的零假设. 所以这个单边符号检验有时用来检验

$$H_0: E(X_i) \geqslant E(Y_i) \quad 对于所有的 i$$

对备择假设：

$$H_1: E(X_i) < E(Y_i) \quad 对于所有的 i$$

对于中位数也有类似的假设表述.

计算机辅助 *Minitab* 和 *StatXact* 可以进行符号检验. ◀

例 3.4.1

物品 A 通过某种过程制成，物品 B 和 A 有同样的功能，但是由一个新过程制成. 制造商想知道物品 B 是否更受消费者欢迎，所以她抽取了由 10 个消费者组成的随机样本，给他们每人一个 A 和一个 B，让他们用一段时间. 符号检验（单边）用来检验

$$H_0: P(+) \leqslant P(-)$$

对备择假设：

$$H_1: P(+) > P(-)$$

其中，“+”代表事件“B 比 A 受欢迎”，“-”代表事件“A 比 B 受欢迎”. 换句话说，H_0 为“B 不比 A 倾向受欢迎”，H_1 为“B 比 A 倾向受欢迎”. 检验统计量 T 是“+”号的个数，即喜欢 B 消费者的人数，临界域对应着 T 值大于等于 $n-t$. 但是，在求 n 和 t 值之前我们需要知道有多少个结点.

在给定的使用时间结束后，消费者给出他们对物品的喜好，8 个消费者喜欢 B，1 个喜欢 A，其余认为“没有差别”. 所以

8 =“+”的个数　1 =“-”的个数　1 = 结点的个数

n =“+”+“-”的个数 = 8 + 1 = 9　T =“+”的个数 = 8 161

对 $n=9$，在表 A3 中用 $p=1/2$，查找近似等于 0.05 的值，则水平为 $\alpha_1 = 0.0195$ 的临界域对应着 T 值大于等于

$$n - t = 9 - 1 = 8$$

因为 $T=8$，所以拒绝 H_0. p-值为 $P(Y \geqslant 8) = 0.0195$.

制造商得出的结论是消费者喜欢 B. ■

在下面的例子中，我们将说明双边符号检验中大样本逼近的使用.

例 3.4.2

Arbuthnott(1710)可能是第一个公开出版的非参数检验的报告，它考查了伦敦 82 年来每年的出生记录，并比较了每年出生的男性和女性的数目. 如果对于每一年我们用“+”代表“出生的男性比女性多”，反之用“-”表示（没有结点），则所考虑的假设为：

$$H_0: P(+) = P(-)$$
$$H_1: P(+) \neq P(-)$$

检验统计量 T 等于“+”号的个数，$\alpha = 0.05$ 水平的临界域对应着 T 值小于

$$t = 0.5(82 - (1.960)\sqrt{82}) = 32.1$$

和 T 值大于

$$n - t = 82 - 32.1 = 49.9$$

其中 t 用（10）式计算.

从记录中，Arbuthnott 得到了 82 个“+”号，没有“-”号，没有“结点”. 所以 $T=82$，并且拒绝零假设. 事实上，可以在更小的 α 水平

$$P(T=0) + P(T=82) = (\tfrac{1}{2})^{82} + (\tfrac{1}{2})^{82} = (\tfrac{1}{2})^{81}$$

下拒绝 H_0，上式是 p-值. ■

为了看到符号检验的更多用途，K. Schmidt-Koenig 向 Batschelet(1965)介绍了下面的例子.

例 3.4.3

把10只信鸽带到它们鸽笼以西25公里的地方，逐个放飞，看它们是随机向各个方向
162 飞（零假设），还是会向朝着它们窝的东方飞. 用固定的野外望远镜观察鸽子直到它们飞离视线，同时记录消失点的角度. 这10个角度为：20,35,350,120,85,345,80,320,280,85度. 令"+"表示更偏东的方向（0到90度或270到360度），"-"表示远离窝的方向（90度和270度之间）. 假设

$$H_0: P(+) \leqslant P(-)$$
$$H_1: P(+) > P(-)$$

正是C部分的右边检验，所以临界域对应着大的 T 值，其中 T 为"+"的个数. 在A3中用 $p=1/2$ 和 $n=10, \alpha=0.0547$ 水平的临界域对应着的 T 值大于等于 $10-2=8$.

对这些数据，可得 $T=9$，所以拒绝零假设. 结果是这些信鸽趋向于飞回家而不是随机地乱飞，p-值是 $P(T \geqslant 9)=0.0107$. ■

□**理论**　事件"+"代表事件"$Y_i > X_i$"，或"$Y_i - X_i > 0$"，即是说差 $Y_i - X_i$ 是正的. 类似地，"-"和"0"分别代表事件 $Y_i - X_i$ 是"负"的或"零". 所以，符号检验是用来比较差为正数的概率和差为负数的概率的检验. 在二项检验中，这些分别被称为"类1"和"类2"概率. 忽略结点，我们有

$$P(+) + P(-) = 1$$

所以假设

$$H_0: P(+) = P(-)$$

这等价说

$$H_0: P(+) = 1/2$$

它与 $p^* = 1/2$ 时，二项检验的形式是一样的. 所以用同样的二项检验方法，从下面的对称性，

$$p^* = 1/2 = 1 - p^*$$

可以得到简化结果. 当符号检验用于检验A,B,C部分的假设时，符号检验是无偏和相合的（见Hemelrijk,1952）. 例2.4.2解释了 $p=1/2$ 的二项检验，和没有结点的符
163 号检验是一样的，所以图2-4中所画的功效函数就是符号检验的功效函数. 虽然没有证明，但从这些图中很明显得出，符号检验是无偏和相合的. □

如果我们对符号检验加上另外的假定，即假设差 $Y_i - X_i$ 是具有对称分布的随机变量［如果对于所有 x，有 $P(Z \leqslant c-x) = P(Z \geqslant c+x)$，则称随机变量 Z 的分布函数关于 c 点对称］，Wilcoxon符号秩检验是更适合的（见5.7节）. 进一步讲，如果差 $Y_i - X_i$ 是独立同分布的正态随机变量，合适的参数检验称为成对 t 检验. 在这种情况下，与成对 t 检验相比，A. R. E. 只为 $2/\pi = 0.637$. 同样与Wilcoxon符号秩检验相比，A. R. E. 为2/3. 如果差有均匀分布（轻尾分布），则符号检验的A. R. E. 相对于 t 检验或Wilcoxon检验则降至 $1/3 = 0.333$. 对于对称的重尾分布，如我们所知的双指数分布，符号检验的A. R. E. 相对于 t 检验和Wilcoxon符号秩检验分别升到2.0

和 4/3 = 1.333.

Walsh(1951),Dixon(1953),Hodges 和 Lehmann(1956)以及 Gibbons(1964)还有其他人研究了大样本和小样本情况的效率. MacKinnon(1964)给出了样本量直到 1000 的特殊表. Hemelrijk(1952)讨论了结点的情况.

像符号检验一样，当数据成对出现时，我们可以通过减少序列对变成单值序列来分析，好像只有一个样本一样，即用单变量方法来分析两变量样本. 在符号检验中，可采用分析一系列值的相同方式来分析差 $Y_i - X_i$，看正值是否比负值多. 需要记住这种降二维（甚至多维）数据为单变量样本的法则，这是很有用的.

习题

1. 6 个学生想通过节食减肥，有如下结果：

姓名	Abdul	Ed	Jim	Max	Phil	Ray
节食之前体重	174	191	188	182	201	188
节食之后体重	165	186	183	178	203	181

   节食是减肥的有效方法吗?

2. 比较了一组 28 名办公室职员的午饭前和午饭后的反应时间，发现 22 名职员午饭前的反应时间更短，2 人没有区别. 午饭后的反应时间显著比午饭前长吗? 164

3. 比较两种不同的添加剂来看哪一种能更好地改进混凝土的耐抗力. 将 100 小批的混凝土在不同条件下混合，混合时每批分为两部分，一部分要加入添加剂 A，另一部分则加入 B. 等到混凝土变硬后，每批的两部分互相挤压，由一个观察员决定哪一部分更耐抗. 在 77 种情况下 A 相对更耐抗；23 种情况下 B 相对更耐抗. 问两种添加剂的耐抗效果有显著差异吗?

4. 邀请在杂货店的 22 名顾客品尝两种干酪并选出喜欢的品种. 7 名顾客喜欢其中的一种，12 名喜欢另一种，3 名没有特别的偏好. 这能说明顾客有明显的偏好吗?

5. 一名妇产科医生认为晚上（下午 6 点到早晨 6 点）比白天出生的婴儿多，但是他的一位统计学家朋友说这只是一种可能. 第二年他们跟踪记录了所有在这个医生照顾下自然生产婴儿的时间，来看谁是对的. 结果是

午夜至凌晨 3 点——16 例	中午至下午 3 点——10 例
凌晨 3 点至 6 点——17 例	下午 3 点至 6 点——11 例
早上 6 点至 9 点——12 例	下午 6 点至 9 点——12 例
上午 9 点至中午——9 例	晚上 9 点至午夜——15 例

   问统计学家是对的吗?

6. 在实验室里，把某种昆虫放在一个平桌子中间所画的圆圈内，在桌子的一边放一种香水吸引这种昆虫. 逐个释放每只昆虫，在一定时间内观察它们是否穿过圆圈的边，如果穿过圆圈的边，则记录下昆虫是朝香气方向的半圆走还是朝远离香气的半圆走. 在合理的时间内，试验结果是：61 只昆虫中有 33 只向香气的方向走，16 只向远离香气的方向走，12 只没有穿过圆圈. 问香水吸引这些昆虫了吗?

思考题

如果在水平 $\alpha=0.05$ 的双边检验中用正态逼近，则 t 值可以用（1）式近似计算出：

$$t_1=\tfrac{1}{2}(n-1.9600\sqrt{n})$$

或近似为

$$t_2=\tfrac{1}{2}n-\sqrt{n}$$

例如，如果 $n=21, t_1=6.009, t_2=5.917$，所以，第一个临界域包含整数 6，但是第二个则不包含整数 6，两个等式得出了不同的临界域. 对 20 到 30 内的哪些 n，两个等式能得出
165 同样的检验？这两个结果是否都等价于 $n=16$ 的情形？

3.5　符号检验的一些变形

假定符号检验中的数据不是有序的，而是具有分别称为“0”和“1”两种类别的名义数据，即每个 X_i 或者是 0 或者是 1，每个 Y_i 也一样. 那么有时要问“我们能发现（0,1）的概率和（1,0）的概率之间的差异吗?”当 X_i 在（X_i,Y_i）中代表试验以前的科目条件（或状态），Y_i 代表同样科目试验后的条件时，这样的问题就出现了. 在符号检验中使用的方法这里同样可以使用，只不过检验的名字有所不同.

▶改变显著性的 McNemar 检验

数据　数据由 n' 个独立的二维随机变量（X_i,Y_i），$i=1,2,\cdots,n'$ 组成. 这里 X_i 和 Y_i 的度量尺度是名义的，分别具有“0”,“1”两个类别；即（X_i,Y_i）的可能值为(0,0)，(0,1)，(1,0)和(1,1). 在 McNemar 检验中，数据常归结为如下的 2×2 的列联表(contingency table).

		Y_i 的分类	
		$Y_i=0$	$Y_i=1$
X_i 的分类	$X_i=0$	a（$X_i=0$ 和 $Y_i=0$ 的对数）	b（$X_i=0$ 和 $Y_i=1$ 的对数）
	$X_i=1$	c（$X_i=1$ 和 $Y_i=0$ 的对数）	d（$X_i=1$ 和 $Y_i=1$ 的对数）

假定条件

1. 数对（X_i,Y_i）是相互独立的.
2. 对所有 X_i 和 Y_i，度量尺度是具有两个类别的名义尺度.
3. 差值 $P(X_i=0,Y_i=1)-P(X_i=1,Y_i=0)$ 或者对于所有 i 为负，或者对于所有 i 为零，或者对于所有 i 为正.

检验统计量　McNemar 检验统计量通常写为

$$T_1 = \frac{(b-c)^2}{b+c} \tag{1}$$

但是，对于 $b+c \leqslant 20$，则选用下面的统计量

$$T_2 = b \tag{2}$$ 166

注意，T_1, T_2 都不依赖于 a 或 d，这是因为 a 和 d 代表“结点”的个数，而这里的分析不考虑结点.

零分布 当 $(b+c)$ 很大时，T_1 的零分布近似于自由度为 1 的 χ^2 分布，T_2 的精确分布是 $p=1/2, n=b+c$ 的二项分布.

假设

$$H_0: P(X_i=0, Y_i=1) = P(X_i=1, Y_i=0) \quad \text{对于所有的 } i$$
$$H_1: P(X_i=0, Y_i=1) \neq P(X_i=1, Y_i=0) \quad \text{对于所有的 } i$$

如果我们将 $P(X_i=0, Y_i=0)$ 加到 H_0 中等式的两边，这些假设的形式会有所不同，即

$$H_0: P(X_i=0, Y_i=1) + P(X_i=0, Y_i=0) = P(X_i=1, Y_i=0) + P(X_i=0, Y_i=0)$$

H_0 的左边包括所有 Y_i 的概率，因此等于 $P(X_i=0)$. 同样，H_0 的右边包括所有 X_i 的概率，所以等于 $P(Y_i=0)$. 从而我们有如下新形式的假设

$$H_0: P(X_i=0) = P(Y_i=0) \quad \text{对于所有的 } i$$
$$H_1: P(X_i=0) \neq P(Y_i=0) \quad \text{对于所有的 } i$$

当然，这也等价于

$$H_0: P(X_i=1) = P(Y_i=1) \quad \text{对于所有的 } i$$
$$H_1: P(X_i=1) \neq P(Y_i=1) \quad \text{对于所有的 } i$$

而后面的假设形式在试验中更易于解释.

令 n 等于 $b+c$. 如果 $n \leqslant 20$，用表 A3. 如果 α 是理想的显著性水平，在表 A3 中用 $n=b+c$ 和 $p=1/2$ 找到近似等于 $\alpha/2$ 的值，称这个值为 α_1，相应的 y 称为 t. 如果 $T_2 \leqslant t$，或 $T_2 \geqslant n-t$，则以 $2\alpha_1$ 的显著水平拒绝 H_0，否则接受 H_0. p-值是 2 倍于 T_2 小于等于观测值的概率和 T_2 大于等于观测值的概率中的较小者，其概率值可在表 A3 中用 $p=1/2, n=b+c$ 获得.

如果 n 超过 20，用 T_1 和表 A2. 所以如果 T_1 超过自由度为 1 的 χ^2 分布的 $(1-\alpha)$ 分位数，则以 α 的显著性水平拒绝 H_0，否则接受 H_0. p-值是 T_1 大于观测值的概率，它可利用自由度为 1 的 χ^2 分布在表 A2 中获得. 更严格的 p-值可以通过比较 T_1 的负平方 167
根和表 A1 而得到，并且左边概率变成了 2 倍.

计算机辅助 *S-Pluss*，*StatXact* 和 *SAS* 可以进行 McNemar 检验. ◀

例 3.5.1

在两个总统候选人的全国电视辩论之前，抽取了一个有 100 人的随机样本，他们对候选人的选择如下，84 个选民主党候选人，剩余的 16 个选共和党候选人. 在辩论之后，又收集了这 100 个人的选择，之前选择民主党候选人的人中有1/4改变了主意，同样之前选共和党的人有 1/4 改选民主党候选人. 结果总结为如下的 2×2 的列联表.

		之后 民主党	之后 共和党	合计 之前
之前	民主党	63	21	84
	共和党	4	12	16
				100

McNemar 检验可以用来检验 H_0：选民总体不受辩论影响，备择假设 H_1：在选民中选民主党人的比例会有变化. 如果第 i 个人之前选民主党，就认为（X_i, Y_i）中的 X_i 为 0；如果选共和党就认为 X_i 为 1. 类似地，Y_i 代表第 i 个人在辩论之后的选择（我们是否选择用 0 或 1 代表对民主党的选择不影响结果，只要 X_i 和 Y_i 用同样的表达即可）. McNemar 检验统计量 T_1 为

$$T_1 = \frac{(b-c)^2}{b+c} = \frac{(21-4)^2}{21+4} = \frac{289}{25} = 11.56 \tag{3}$$

$\alpha = 0.05$ 水平的临界域对应着所有 T_1 值大于 3.841（自由度为 1 的 χ^2 分布的 0.95 分
168 位数，可从表 A2 中查到）. 因为 11.56 大于 3.841，所以拒绝零假设，结论是选民的队伍有所改变，且 p-值小于 0.001. ■

□**理论** 这是一个符号检验的变形，其中事件（0,1）称为"+"，事件（1,0）称为"-"，事件（1,1）和（0,0）称为结点. 则 McNemar 假设检验的形式取为

$$H_0: P(+) = P(-)$$

这和双边符号检验时的 H_0 一样. T_2 的临界域与符号检验 $n \leqslant 20$ 的情况一样.

对于 n 大于 20，建议用正态逼近的符号检验，这是因为当 H_0 为真时，如下表达式

$$Z = \frac{T_2 - n(\frac{1}{2})}{\sqrt{n(\frac{1}{2})(\frac{1}{2})}} = \frac{b - n(\frac{1}{2})}{(\frac{1}{2})\sqrt{n}} \tag{4}$$

近似地服从标准正态分布（见习题 1.5.6）. 因为 $n = b + c$，（4）式化简为

$$Z = \frac{b - [(b+c)/2]}{(\frac{1}{2})\sqrt{b+c}} = \frac{b-c}{\sqrt{b+c}} \tag{5}$$

所以

$$T_1 = Z^2$$

近似地服从自由度为 1 的 χ^2 分布（见定理 1.5.3）. 一个包括 T_2 或 Z 的双边检验的临界域可以与用 $T_1 = Z^2$ 的上尾作为临界域相比. □

由于已经介绍了符号检验的单边和双边形式，所以 McNemar 检验也有这两种形

式. 给出单边形式 McNemar 检验的最简单方法就是用单边符号检验. Bennett 和 Underwood(1970), Ury(1975), Mantel 和 Fleiss(1975), 和 Mckinlay(1975)讨论了 McNemar检验和它的变形.

Cox 和 Stuart(1955)介绍了符号检验的另一种修正形式, 用它来检验某种趋势(trend)的出现. 一列数如果后面的数比前面的数趋于变大(上升趋势)或比前面的数趋于变小(下降趋势), 则称为是有趋势的. 这个检验将后面的数和前面的数组成对, 169
并在所形成的对上进行符号检验. 如果有趋势, 则每对中的一个数比另一个数有变大或变小的趋势; 另一方面, 如果没有趋势, 这列数实际上代表独立同分布的随机变量的观测, 每一对中的任一个数都没有超过另一个的趋势.

▶Cox 和 Stuart 趋势性检验

数据 数据由随机变量序列 $X_1, X_2, \cdots, X_n$ 的观测值组成, 以某种顺序排列, 例如观测的顺序. 我们希望知道这个数列中是否有趋势存在. 把随机变量进行配对分组$(X_1, X_{1+c}), (X_2, X_{2+c}), \cdots, (X_{n'-c}, X_{n'})$, 其中如果 n' 是偶数, 则 $c = n'/2$, 如果 n' 是奇数, 则 $c = (n'+1)/2$. (注意用这种方案时, 如果 n' 是奇数要除去中间的随机变量). 如果 $X_i < X_{i+c}$, 则用"+"代替(X_i, X_{i+c}), 如果 $X_i > X_{i+c}$, 则用"-"代替(X_i, X_{i+c}), 而不考虑结点. 没有结点对的个数称为 n.

这个检验可以用来检测任何给定的非随机模式, 如正弦波或其他周期模式. 随机变量列只是重新排列从而使得最小的数(像预期的那样)靠近数列的开始, 较大的数靠近数列末尾. 则排列后数列出现上升趋势预示着原数列中有某种预期的模式.

假定条件

1. 随机变量 $X_1, X_2, \cdots, X_{n'}$ 是互相独立的.
2. X_i 的度量尺度至少是须序的.
3. X_i 是同分布或有某种趋势; 即后面的随机变量更可能比前面的大(或反之亦然).

检验统计量 $T =$"+"的个数

零分布 统计量的零分布是 $p = 1/2, n =$ 没有结点对数的二项分布, 其中 X_i 不等于 X_{i+c}.

假设 检验的其他部分和前面所述的符号检验是一样的, 这里不重述了. 零假设: 没有出现趋势. 右边单边检验可用于检测上升趋势; 左边检验可用于检测下降趋势; 而双边检验可用于检测存在任何(上升或下降)趋势的备择假设. ◀ 170

下面是用到双边 Cox 和 Stuart 趋势性检验的例子.

例 3.5.2

有 19 年的每年降水量的记录. 检查这个记录来看降水量是否有增加或减少的趋势. 降水量(以英寸为单位)分别是 45.25, 45.83, 41.77, 36.26, 45.37, 52.25, 35.37, 57.16, 35.37, 58.32, 41.05, 33.72, 45.73, 37.90, 41.72, 36.07, 49.83, 36.24 和 39.90.

因为 $n'=19$ 是奇数，去掉中间的数 58.32，把剩下的数组成对.

(45.25, 41.05)	(45.37, 41.72)	(4.5.83,33.72)
(52.25, 36.07)	(41.77, 45.73)	(3.5.37, 49.83)
(36.26, 37.90)	(57.16, 36.24)	(35.37, 39.90)

没有结点，所以 $n=9$. 检验统计量 T 等于第二个数大于第一个数的数对个数. 0.0390 水平的临界域对应着 T 小于等于 1 和 T 大于等于 $9-1=8$ 的值.

代入数据，得到 $T=4$，它在接受域中，且 p-值等于 1.0. 所以接受零假设“没有趋势存在”. ■

在例 3.5.2 中，对于检验为有效的模型假设是合理的，所以检验是合理有效的. 但是，所列假设不都是必要的，我们的假设只需能满足符号检验的模型就足够了. 即只需要假设：

1. 随机变量 $X_1, X_2, \cdots, X_n$，是互相独立的.
2. 对于所有的对，概率 $P(X_i < X_{i+c})$ 和概率 $P(X_i > X_{i+c})$ 有同样的相对大小.
3. 每对 (X_i, X_{i+c}) 能够判别为“+”，“-”或“结点”.

这些假设不像检验中所给的系列假设那样容易理解，但是在很多实际问题中它们更有用，例如下面的例子.

例 3.5.3

24 个月中，记录了某条小溪每月的平均水流速度（单位：立方英尺/秒）. 要检验的假设是：

H_0: **平均水流速度没有降低**

备择假设：

171 H_1: **平均水流速度降低了**

我们知道水流速度是以年为周期的，所以将两个不同月的水流速度配对是无济于事的. 但是，将连续两年的同一个月配对，这样可以进行趋势性研究. 收集数据如下：

月份	第一年	第二年	月份	第一年	第二年
一月	14.6	14.2	七月	92.8	88.1
二月	12.2	10.5	八月	74.4	80.0
三月	104	123	九月	75.4	75.6
四月	220	190	十月	51.7	48.8
五月	110	138	十一月	29.3	27.1
六月	86.0	98.1	十二月	16.0	15.7

检验统计量 T 等于第二年流速比第一年高的对数，本例中 $T=5$. 因为是检验下降趋势，0.0730 水平的临界域对应着所有 T 小于等于 3（从表 A3 中，用 $n=12, p=1/2$ 获得的值）. 所以接受 H_0，p-值为

$$P(T \leqslant 5 \mid H_0 \text{ 成立}) = 0.3872$$

这个值太大，以至于是一个不能接受的 α. ■

这一节中的例子只能代表符号检验所适用的不同假设检验的一小部分. 再

用两个例子来结束这一节. 对第一个例子符号检验作为检验相关性的一个简单方法，即检验一个随机变量的较大值是否趋向于和第二个随机变量的较大值配对，而较小值和较小值配对（正相关），或一个随机变量的较大值是否趋向于和第二个随机变量的较小值配对，而较小值和较大值配对（负相关）. 检验包括排数对（数对伴随不变）使得数对中的一个数（通常是结点较少的变量，第一个或第二个变量）的顺序是递增的. 如果有相关性，数对中的另一个数将会呈现出趋势性，如果是正相关就是上升趋势，如果是负相关就是下降趋势. Cox 和 Stuart 趋势性检验就可用于由数对中另外一个数形成的数列上.

例 3.5.4

Cochran(1937)比较了一些病人对两种药的反应，来说明每个病人对两种药的反应是否有正相关性. 172

病人	药物 1	药物 2	病人	药物 1	药物 2
1	+0.7	+1.9	6	+3.4	+4.4
2	−1.6	+0.8	7	+3.7	+5.5
3	−0.2	+1.1	8	+0.8	+1.6
4	−1.2	+0.1	9	0.0	+4.6
5	−0.1	−0.1	10	+2.0	+3.4

根据对第一种药物的反应对数对排序得到：

病人	药物 1	药物 2	病人	药物 1	药物 2
2	−1.6	+0.8	1	+0.7	+1.9
4	−1.2	+0.1	8	+0.8	+1.6
3	−0.2	+1.1	10	+2.0	+3.4
5	−0.1	−0.1	6	+3.4	+4.4
9	0.0	+4.6	7	+3.7	+5.5

把单边的 Cox 和 Stuart 趋势性检验用于新排序后药物 2 的数列上. 产生了 5 个数对为：(+0.8, +1.9), (+0.1, +1.6), (+1.1, +3.4), (−0.1, +4.4), (+4.6, +5.5). 因为我们检验：

$$H_0\text{: 没有正相关性}$$

备择假设：

$$H_1\text{: 有正相关性}$$

本质上，我们要检验上升趋势（H_1），检验统计量 T 等于 5，因为在所有 5 对数据中药物 2 的第二个观测超过了第一个观测. 0.0312（对 $n=5, p=1/2$ 用表 A3 中，并得 $t=0$）水平的临界域对应于单值 $T=5$，所以拒绝零假设，从而可得出对两个药物的反应有正相关性. 本例中的 p-值也等于 0.0312. ■

最后的一个例子用于解释符号检验或 Cox 和 Stuart 趋势性检验怎样检验预期的模式.

例 3.5.5

在一个 24 小时的试验中，以小时为单位记录实验室中的一群昆虫产卵的数量，要检验

173 H_0: 24 个产卵数量组成 24 个同分布随机变量的观测值.

备择假设

H_1: 产卵数量在下午 2:15 达到最小，逐渐增加直到凌晨 2:15 增大到最大值，再减少直到下午 2:15.

每小时产卵数量的记录如下

时间	卵的数量	时间	卵的数量	时间	卵的数量
上午 9 点	151	下午 5 点	83	凌晨 1 点	286
上午 10 点	119	晚上 6 点	166	凌晨 2 点	235
上午 11 点	146	晚上 7 点	143	凌晨 3 点	223
中午 12 点	111	晚上 8 点	116	凌晨 4 点	176
下午 1 点	63	晚上 9 点	163	凌晨 5 点	176
下午 2 点	84	晚上 10 点	208	早上 6 点	174
下午 3 点	60	晚上 11 点	283	上午 7 点	139
下午 4 点	109	晚上 12 点	296	上午 8 点	137

如果备择假设成立，卵的数量在离下午 2:15 最近时应当趋于最少，凌晨 2:15 附近应当趋于最多. 所以，根据时间从下午 2:15 左右到凌晨 2:15 左右重新排列卵的数量.

时间	卵的数量	时间	卵的数量
下午 2 点	84	上午 8 点	137
下午 3 点	60	晚上 9 点	163
下午 1 点	63	上午 7 点	139
下午 4 点	109	晚上 10 点	208
中午 12 点	111	早上 6 点	174
下午 5 点	83	晚上 11 点	283
上午 11 点	146	凌晨 5 点	176
晚上 6 点	166	晚上 12 点	296
上午 10 点	119	凌晨 4 点	176
晚上 7 点	143	凌晨 1 点	286
上午 9 点	151	凌晨 3 点	223
晚上 8 点	116	凌晨 2 点	235

如果 H_1 成立，这些数据应当呈现出上升趋势. 用单边 Cox 和 Stuart 趋势性检验，用数列的前一半（第一列）与数列的后一半（第二列）配对，则每一行上的产卵数形成一个数对. 在全部的 12 对中，第二列的数都大于第一列，所以 $T=12$. 对于 $n=12, p=1/2$，在表 A3 中查出 $\alpha=0.0193$ 水平的临界域对应着 T 值大于等于 $12-2=10$. 所以拒
174 绝 H_0，我们可以得出确实有我们预期的模式. p-值由下式给出：

$$P(T \geqslant 12)=0.0002$$

所以，我们可以以任何合理的水平拒绝 H_0. ■

□**理论** Cox 和 Stuart 趋势性检验显然是符号检验的变形，所以当 H_0 成立时，检验统计量的分布明显是二项分布. 而且用前面的第一部分 A,B,C 假设检验时，这个检验是无偏和相合的，但用后一部分时，则未必是. 当用于正态随机变量时，Stuart (1956) 证明这个检验关于最好参数检验（基于回归系数的检验）的 A. R. E. 是 0.78，在同样的情况下，对于 Spearman 或 Kendall 的秩相关性检验（将在第 5 章介绍的随机性检验）的 A. R. E. 是 0.79.

如果检验变为除去中间 1/3 的观测，将观测前面的 1/3 和后 1/3 配对，则在理想条件下，它关于参数检验的 A. R. E. 会增到 0.83. 显然，与获得的较大偏差相比，数据的损失是小的，这表明另一种变化，即从数列的两端进行配对，用所有的数据组成(X_1,X_n),(X_2,X_{n-1})等，这可能保留了大的偏差，而数据没有丢失. 同上面过程一样可以进行检验，因为在零假设下，检验统计量的分布没有发生变化.

对于在例 3.5.4 中给出的相关性检验，我们没有研究它具有什么性质. 应用相关性检验的一个困难是如果很多观测值相等，因此，就有不止一种用于趋势性检验的观测排列方法，所以，推荐用有最少结点的对数来排列原始数据. 因为排列可能还有结点，保守的方法是选择最不可能拒绝 H_0 的排列. □

Chatterjee(1966) 讨论了位置参数的两维的符号检验. 符号检验的另外一个变形就是用于散布的趋势性检验（Ury,1966），或比较带有一个控制的几个处理（Rhyne 和 Steel,1965）. Rao(1968) 把 Cox 和 Stuart 检验用于散布的趋势性检验. Mansfield (1962) 进一步讨论了趋势性检验的功效. Olshen(1967) 提出了二次趋势对线性趋势的检验. Woodbury,Manton,Woodbury(1977) 和 Altham(1971) 给出了符号检验的其他形式. Schaafsma(1973) 的论文检验了符号检验次序依赖的结果，即，一个顾客喜欢 175
哪一种品牌可能受他（或她）第一次所接触品牌的影响.

习题

1. 随机抽取 135 名美国公民，问他们对美国外交政策的看法，43 人反对美国的外交政策. 在之后的几周内他们收到了一份时事信息报，然后再问他们的观点；有 37 人反对，而 37 人中的 30 人是之前不反对美国外交政策的人. 问反对美国外交政策的人数有显著变化吗?
2. 在习题 1 中，假设试验之后反对美国外交政策的 37 个人是之前也反对的人. 问反对美国外交政策的人数有显著变化吗?
3. 在某个城市，最近 15 年内每 100,000 个人中交通事故的死亡率分别是：17.3,17.9,18.4,18.1,18.3,19.6,18.6,19.2,17.7,20.0,19.0,18.8,19.3,20.2,19.9. 认为死亡率在增加是否有根据?
4. 中西部某所小型的大学在最近 34 年记录了每年大一男生的身高. 平均数为 68.3,68.6,68.4,68.1,68.4,68.2,68.7,68.9,69.0,68.8,69.0,68.6,69.2,69.2,68.9,68.6,68.6,68.8,69.2,68.8,68.7,69.5,68.7,68.8,69.4,69.3,69.3,69.5,69.5,69.0,69.2,69.2,69.1,69.9. 问这些数据有上升趋势吗?
5. 某制造商计算了 44 个月生产某种产品所需费用的平均值（以美元为单位）：13.65,13.41,

13.53,13.23,13.58,13.43,13.73,13.40,13.70,13.58,13.80,13.40,13.63,13.69,13.92,13.68,13.72,13.42,13.66,13.98,13.81,13.60,13.32,13.45,13.27,13.26,13.28,13.29,13.10,13.09,13.36,13.40,13.35,13.53,13.66,13.10,13.28,13.33,13.02,13.09,13.12,13.16,12.96,12.95. 问这些平均数是否有统计意义下的趋势?

6. 在一个研究他人意见对人有无影响的试验中，把不同长度的 20 根线每次一根分别放在被测者 A 和 B 面前，要求他们大声地估计出每根线的长度. 在被测者 B 不知道的情况下，指导被测者 A 先说出她的估计，让她高估前 10 根，低估后 10 根，在听了 A 的估计后，B 再说出他的估计值. 估计的错误是估计值减去真实值的差，记录如下

	线									
	1	2	3	4	5	6	7	8	9	10
A的错误	+0.3	+1.1	+0.9	+0.6	+1.0	+1.3	+0.8	+1.6	+1.2	+0.8
B的错误	−0.1	+0.6	+1.0	+0.7	+0.2	+0.9	−0.1	+0.2	0.0	+0.5

176

	线									
	11	12	13	14	15	16	17	18	19	20
A的错误	−1.3	−1.1	−1.3	−0.7	−1.4	−1.1	−0.8	−0.5	−1.2	−1.0
B的错误	−0.6	−1.2	−1.0	−0.7	−1.0	−0.1	−0.5	0.0	−0.4	−0.3

问被测者 A 和 B 的错误之间是否有明显的正相关?

7. 下面是一名职业棒球队主力队员 12 年中本垒打次数和平均击球数的记录.

	1988	1989	1990	1991	1992	1993
本垒打次数	7	14	17	15	9	19
平均击球数	0.212	0.232	0.234	0.210	0.201	0.256

	1994	1995	1996	1997	1998	1999
本垒打次数	16	17	22	17	13	10
平均击球数	0.261	0.247	0.255	0.241	0.238	0.235

问他每年本垒打的次数与平均击球次数是否有明显的相关性?

8. 检验下面的数据看一个家庭的年收入和该家庭中孩子的个数是否有显著的相关性.

收入（美元）	孩子个数	收入	孩子个数	收入	孩子个数
17,440	3	23,320	3	28,940	3
17,664	2	23,569	4	29,300	1
17,721	4	23,950	2	29,371	3
17,883	3	24,023	3	29,512	1
18,000	4	24,330	5	29,662	1
18,332	2	24,545	2	29,804	2
18,653	0	24,922	5	30,167	2
18,781	3	25,571	4	30,634	3
19,087	6	25,624	4	31,235	1
19,686	5	25,873	2	31,797	3
19,832	2	26,010	1	31,880	4
20,100	1	26,145	3	32,363	1
20,222	6	26,513	2	32,946	3
20,435	3	26,660	4	33,586	2

收入（美元）	孩子个数	收入	孩子个数	收入	孩子个数
20,961	5	26,984	5	34,000	2
21,382	2	27,463	0	34,443	3
21,957	0	27,702	1	35,693	1
22,190	8	27,914	4	39,247	1
22,212	1	28,244	2	40,540	1
22,635	4	28,698	4	55,686	2

9. 对相距 50 英里的两个飞机场进行一年的观测，来断定天气条件对可用性是否有显著差异. 有 286 天两个机场都全天开放，有 62 天由于严峻的天气两个机场至少都关闭一段时间，有 14 天机场 A 关闭但 B 开放，3 天相反. 问由于天气条件的可用性是有显著的差异吗？ 177

思考题

1. 某理发店正考虑将理发的价格提高 1 美元，同时给顾客一张可以在附近一家酒吧喝饮料的免费优惠券. 进行一项调查，在实际顾客和潜在顾客（非顾客）中随机选择 200 个作为样本，告诉他们这项提议. 样本中 10% 的顾客表示他们会到别的理发店去理发，样本中非顾客的 5% 表示他们会到这里来理发. 检验零假设：如果样本中只有 20 个是目前的顾客，提出的建议不会增加来这个理发店理发的人数. 如果目前的顾客是 60，结果会有怎样的变化？
2. McNemar 检验的数据可以写成二维观测 X_i, Y_i，其中每个观测为 0 或 1 “之前”，0 或 1 “之后”. 称为“成对 t 检验”的参数检验常常用于这种类型的数据，成对 t 检验使用差 $D_i = X_i - Y_i, i = 1, 2, \cdots, n'$. 在检验统计量中使用样本均值 $\overline{D}$ 和样本标准差 S：

$$t = \overline{D}\sqrt{n'-1}/S$$

将 t 值与自由度为 $n'-1$ 的 t 分布的分位数（在表 A21 中可获得）相比较，如果 D_i 不是正态分布，则这个检验只是一个近似.

证明 t 和 T_1 的如下关系成立：

$$t^2 = \frac{(n'-1)T_1}{n'-T_1} \qquad 或 \qquad T_1 = \frac{n't^2}{n'-1+t^2}$$

其中，T_1 由（1）式给出. 即当 T_1 变大时，t^2 也变大，所以，如果它们的临界域相互对应（当 T_1 较大或 t^2 较大时，拒绝 H_0），则这两个检验是等价的. 178

第4章 列 联 表

导 言

列联表（contingency table）是一列按照矩阵形式排列的自然数，这些自然数通常代表的是数量或者频数．例如：昆虫学家可以说他正在观察37只昆虫，也可以用1×3列联表描述他所观察到的：

月	蚱蜢	其他	总和
12	22	3	37

因为只有一行，所以这个列联表是一维的.

这个昆虫学家可能希望更具体一些，于是他用了一个2×3列联表来描述：

	月	蚱蜢	其他	总和
活着	3	21	3	27
死亡	9	1	0	10
总和	12	22	3	37

这里的总和包括两个*行总和*（row total），三个*列总和*（column total）与一个所有
179 行列总和，它们是可选择列在表上的，通常只是为了读者的方便．这个列联表是二维的，并且可以扩展到含有r行和c列的$r\times c$列联表．三维及三维以上的列联表也有可能出现，但本章只对它们做简要的讨论.

4.1 2×2列联表

一般$r\times c$列联表是一排排成r行c列的自然数，因此有rc个数格或者位置来放置这些数．本节主要讨论$r=2$和$c=2$的情形，即2×2列联表，因为包含四个数格，所以2×2列联表也被称为*四重*（fourfold）列联表.

从某个总体中随机选取的N个对象，在处理或者一个事件发生前将它们归入两类中的某一类，这时可以用2×2列联表．在处理后再次检查这N个对象并且分成两类．需要解决的问题是：这种处理是否明显改变了每类中所含对象的比例？列联表的使用将在3.5节中介绍．可以看到，合适的统计方法是符号检验的变形，即McNemar检验．因为相同的样本用在两种情形中（比如在“处理前”和“处理后”），所以McNemar检验能够发现微小的差异．同一假设检验对应于McNemar检验的另一检验方法就是从处理前后的总体中各抽取一随机样本，然后做比较．但是这种使用两个不同的随机样本

又带来了另外不理想的变异，因为这种变异使得在总体中由处理所引起的变化不明显．然而，在不实用或甚至是可能使用同样的样本两次时，那么可使用本节将要描述的方法．

对于零假设：某事件 A（某些特定事件）在两个总体中发生的概率相同（零假设也可以表述为：具有特征 A 的总体比例对两个总体是相同的）．第一步需要从两个总体中各抽取一个随机变量来检验原假设．

▶2×2 概率差异的 χ^2 检验

数据 从一个总体（或处理前）中抽取一个具有 n_1 个观测的随机样本，并且每个观测被归入类 1 或类 2，两类的观测数分别是 O_{11} 和 O_{12}，并且 $O_{11}+O_{12}=n_1$．从另一个总体(或者处理后的第一个总体)中抽取 n_2 个观测并且记类 1 与类 2 中的样本数分别为 O_{21},O_{22}，这里 $O_{21}+O_{22}=n_2$．数据可以被放在下面的 2×2 列联表中，观测总数记为 N． 180

	类 1	类 2	总和
总体 1	O_{11}	O_{12}	n_1
总体 2	O_{21}	O_{22}	n_2
总和	C_1	C_2	$N=n_1+n_2$

假定条件

1. 每个样本是随机样本．
2. 两个样本相互独立．
3. 每一个观测可以被归入类 1 和类 2 中的任一个．

检验统计量 如果任意列总和是 0，那么检验统计量 $T_1=0$，否则，

$$T_1=\frac{\sqrt{N}(O_{11}O_{22}-O_{12}O_{21})}{\sqrt{n_1n_2C_1C_2}} \tag{1}$$

零分布 因为所有的 O_{11},O_{12},O_{21} 和 O_{22} 可能值有不同的结合，所以 T_1 的精确分布难以用表的形式表示出来．因此我们用大样本逼近或标准正态分布逼近，其分位数在表 A1 中给出．

假设 从总体 1 中随机抽取的观测归入类 1 的概率记为 p_1，从总体 2 中随机抽取的观测归入类 1 的概率记为 p_2，这里 p_1 和 p_2 不必已知，且假设仅仅是给定它们之间的关系．

A.（双边检验）

$$H_0: p_1=p_2$$
$$H_1: p_1\neq p_2$$ 181

α 为近似水平，如果 T_1 比标准正态随机变量 Z 的 $\alpha/2$ 分位数小或者 T_1 比 Z 的 $1-\alpha/2$ 分位数大，则拒绝 H_0，Z 的分位数在表 A1 中可查到．

从表 A1 中看出，p-值是 Z 小于 T_1 的观测值或大于 T_1 的观测值的概率中较小者

的2倍.

注意，对上述假设，T_1^2 常代替 T_1 作为统计量，那么拒绝域就是自由度为1的卡方分布的右边，在表A2中可查到.

B.（左边检验）

$$H_0: p_1 \geqslant p_2$$
$$H_1: p_1 < p_2$$

α 为近似水平，如果 T_1 比标准正态随机变量 Z 的 α 分位数小，则拒绝 H_0，Z 的分位数在表A1中可查到.

从表A1中看出，p-值是 Z 小于 T_1 观测值的概率.

C.（右边检验）

$$H_0: p_1 \leqslant p_2$$
$$H_1: p_1 > p_2$$

α 为近似水平，如果 T_1 比标准正态随机变量 Z 的 $1-\alpha$ 分位数大，则拒绝 H_0，Z 的分位数在表A1中可查到.

从表A1中看出，p-值是 Z 大于 T_1 观测值的概率.

计算机辅助 *Minitab*,*S-Plus*,*SAS* 和 *StatXact* 可完成这个检验，并且正如思考题3.1.2中介绍的那样，还能求出两个概率差异的置信区间. ◄

例 4.1.1

从两辆货车上装的产品中随机抽样，来检查两车货物的次品率是否有差异. 第一辆

182 货车的86件产品中有13件次品，第二辆的74件产品中有17件次品.

	次品	非次品	总和
货车 1	13	73	86
货车 2	17	57	74
总和	30	130	160

它满足我们前面的假设条件，所以用双边检验来检验 H_0：两个货车上的次品比例相等用如下检验统计量：

$$T_1 = \frac{\sqrt{N}(O_{11}O_{22} - O_{12}O_{21})}{\sqrt{n_1 n_2 C_1 C_2}}$$
$$= \frac{\sqrt{160}((13)(57) - (73)(17))}{\sqrt{(86)(74)(30)(130)}} = -1.2695$$

从表A1中查到标准正态分布的0.975分位数是1.9600，因此近似大小0.05的拒绝域包含 T_1 大于1.9600或小于-1.9600的所有值. 观测值是-1.2695，所以在显著性水平 $\alpha=0.05$ 下可以接受零假设.

p-值是 Z 小于 T_1 的观测值-1.2695的概率较小者的2倍，在表A1中可以查到，是0.102，所以近似 p-值是0.204，因此接受 H_0 还是相当安全的. ■

以下给出使用单边检验的例子.

例 4.1.2

美国海军学院给学生的住所安装了一种新的照明系统. 要说明新的照明系统会导致视力下降，因为使学生的眼睛处于连续的疲劳中. 考虑研究检验零假设：

H_0:毕业学生在新照明系统下有 20-20（好）视力的概率大于等于在旧照明系统下好视力的概率.

对单边备择假设：

H_1:新系统下好视力的概率小于旧系统下好视力的概率.

令 p_1 为随机选取的毕业学生在旧的照明系统下有好视力的学生概率，p_2 为是新系统下相应的概率，则前面的假设可以表述为： 183

$$H_0: p_1 \leqslant p_2$$
$$H_1: p_1 > p_2$$

这与假设集 C 匹配. 新系统建立前的所有毕业班作为总体 1，4 年使用新灯光系统的第一个毕业班作为总体 2，从这两个总体中随机取样. 希望这些样本同先前全体毕业班总体中得到随机样本中真实的和潜在的表现一样.

假定有如下结果：

	好视力	差视力	
旧系统	$O_{11}=714$	$O_{12}=111$	$n_1=825$
新系统	$O_{21}=662$	$O_{22}=154$	$n_2=816$
总和	1376	265	$N=1641$

判决法则 C 定义了 $\alpha=0.05$ 的临界域是 T_1 大于 1.6449（从表 A1 中获得）的所有值. 计算 T_1 得到：

$$T_1=\frac{\sqrt{N}(O_{11}O_{22}-O_{12}O_{21})}{\sqrt{n_1n_2C_1C_2}}=\frac{\sqrt{1641}[(714)(154)-(111)(662)]}{\sqrt{(825)(816)(1376)(265)}}=2.982$$

所以，显然拒绝零假设. 从表 A1 中我们看到，在显著性水平大约为 0.002 时也应该拒绝零假设，所以 p-值是 0.002.

因此我们可以得出，代表两个毕业班总体差视力的比例的确不同，并且可以预见发展趋势. 即：总体 2（新的光照系统下）的视力比总体 1（在旧系统下）更差. 视力更差是否是新的光照系统导致的还不能说明. 然而，在这个假设检验的例子中可以说明，视力变差的原因同新的光照系统有关. ■

□**理论** 这里所给出的 2×2 列联表是下一节所给出的 $r\times c$ 列联表的一种特殊情况，所以相关的理论也是后面的 $r\times c$ 情形的特殊情况. 然而，除了 r 和 c 很小的情况外，检验统计量的精确分布难以求出. 因此这里我们要给出 T_1 的精确分布. 184

当 $H_0: p_1 = p_2 = p$ 为真时，T_1 的精确概率分布可以如下计算，对于总体 1 中的样本，类 1 中 x_1 项的概率和类 2 中 $n_1 - x_1$ 项的概率通过二项分布给出

$$P\left(\begin{array}{c|c|c} & \text{类 1} & \text{类 2} \\ \text{总体 1} & x_1 & n_1 - x_1 \end{array}\right) = \binom{n_1}{x_1} p^{x_1}(1-p)^{n_1-x_1} \tag{2}$$

类似地，对于总体 2 中抽取的样本，类 1 中 x_2 项的概率和类 2 中 $n_2 - x_2$ 项的概率如下：

$$P\left(\begin{array}{c|c|c} & \text{类 1} & \text{类 2} \\ \text{总体 2} & x_2 & n_2 - x_2 \end{array}\right) = \binom{n_2}{x_2} p^{x_2}(1-p)^{n_2-x_2} \tag{3}$$

因为两个样本是独立的，所以联合事件的概率可以通过（2）和（3）式的右边相乘得到，即：

$$P\left(\begin{array}{c|c|c} & \text{类 1} & \text{类 2} \\ \text{总体 1} & x_1 & n_1 - x_1 \\ \text{总体 2} & x_2 & n_2 - x_2 \end{array}\right) = \binom{n_1}{x_1}\binom{n_2}{x_2} p^{x_1+x_2}(1-p)^{N-x_1-x_2} \tag{4}$$

在简单的情形下，我们取 $n_1 = 2$ 和 $n_2 = 2$，样本空间中有 9 个不同的点，对应有如下 9 个可能的表：

表		如果 H_0 为真的概率 $(p = 1/2)$	$(p = 1)$	T_1
2 0 2 0	p^4	1/16	1	不确定
2 0 1 1	$2p^3(1-p)$	1/8	0	1.1547
2 0 0 2	$p^2(1-p)^2$	1/16	0	2.0000
1 1 2 0	$2p^3(1-p)$	1/8	0	-1.1547
1 1 1 1	$4p^2(1-p)^2$	1/4	0	0
1 1 0 2	$2p(1-p)^3$	1/8	0	1.1547

185

0 2 2 0	$p^2(1-p)^2$	1/16	0	−2.0000
0 2 1 1	$2p(1-p)^3$	1/8	0	−1.1547
0 2 0 2	$(1-p)^4$	1/16	0	不确定

因为0/0未定义，所以 T_1 的值不确定. 但正如第5个结果能很强地指出 H_0 为真一样，导致 T_1 值不确定的两个结果也能很强地指出 T_0 为真，所以同第5个结果保持一致，我们可以定义第一个和最后一个结果中的 T_1 为0. 那么 T_1 有如下概率分布：

$p=1/2$	$p=1$
$P(T_1=-2)=1/16$	$P(T_1=0)=1$
$P(T_1=-1.1547)=1/4$	
$P(T_1=0)=3/8$	
$P(T_1=1.1547)=1/4$	
$P(T_1=2)=1/16$	

类似地，任意大小为 n_1 和 n_2 的样本的精确分布，可以通过适当地定义 T_1 的不确定值来求出. 然而，正如前面的例子所示，即使当 H_0 为真时，概率分布函数也不是唯一的，而是依赖于 p. 因此在前面检验中零假设是一个复合假设. 但不容易看出来的是，当 $p=1/2$ 时，前面小样本情形的临界域水平最大. 因此通过令 $p=1/2$ 可以求出 α. 如果临界域对应于最大的 T_1 值(i.e.，$T_1=2$)，则 $\alpha=0.0625$. 186

为了说明正态分布作为大样本的逼近分布是合理的，考虑 $O_{11}/n_1-O_{21}/n_2$ 的均值和方差分别是 p_1-p_2，$p_1q_1/n_1+p_2q_2/n_2$，而在假设 H_0：$p_1=p_2$ 下，均值是0，方差用估计量代替，其中，用 C_1/N 估计 p，C_2/N 估计 $q=1-p$. 由中心极限定理 O_{11} 和 O_{21} 是渐近正态的，所以 $O_{11}/n_1-O_{21}/n_2$ 也是渐近正态的. 在 H_0 下减去均值（0），除以估计的标准差，我们得到：

$$\frac{O_{11}/n_1-O_{21}/n_2}{\sqrt{\frac{C_1C_2}{NN}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \tag{5}$$

当 H_0 为真时，上式渐近为一标准正态随机变量. 然而（5）式的表达式经过简单的代数运算，它就是检验统计量 T_1，所以说明了 T_1 的零分布是渐近于标准正态. □

当容量为 N 的单个样本的每个观测根据两种属性分类，每种属性可取两种形式之一时，也可以使用2×2列联表. 那么就有（2）（2） = 4种两种属性的不同组合，并且2×2列联表是将每一类中的观测数列成表的最方便的方式. 其实，2×2列联表

的这种用法是 $r \times c$ 列联表的特殊情形，且分开表述也没有任何特殊的变化（比如本节的单边检验）．因此我们在下一节中来阐述它．

列联表的这种类型同第一种类型的主要区别是这种列联表的行总和是随机变量，它的值只有在数据获得后才可确定．而第一种列联表行总和表示了两种样本的样本容量，在数据获得前它的值就已知，所以是非随机的．两种表的列总和都是随机变量．

列联表的第三种类型是行总和与列总和都是非随机的，也就是说，行总和与列总和在数据获得前就已知．这种情况，不如列联表的前两种类型常见，但是不论是哪一种类型，经常会用到以下的统计方法，因为它很容易确定精确的 p-值．
187 这个方法在 20 世纪 30 年代中期几乎同时由 R. A. Fisher(1935)，I. O. Irwin(1935) 和 F. Yates(1934) 发展起来的，而众所周知的是 Fisher 的精确检验．

▶Fisher 精确检验

数据　如同前面的检验，除了行总和 r 和 $N-r$，列总和 c 和 $N-c$ 事先确定（从而不随机）外，把数据中的 N 个观测放在 2×2 列联表中．

	列 1	列 2	
行 1	x	$r-x$	r
行 2	$c-x$	$N-r-c+x$	$N-r$
	c	$N-c$	N

假定条件

1. 每个观测只归入到一个单元中．
2. 行列总和确定且不随机（但注意结尾关于行、列和行列随机总和的注释）．

检验统计量　检验统计量 T_2 是第一行第一列单元格中的观测数．

零分布　当 H_0 为真时，T_2 的精确分布由超几何分布给出（见（1.3.17）式），对 $x=0,1,\cdots,\min(r,c)$

$$P(T_2=x)=\frac{\binom{r}{x}\binom{N-r}{c-x}}{\binom{N}{c}} \qquad x=0,1,\ldots,\min(r,c)$$

$$=0 \qquad \text{对其他 } x. \tag{6}$$

用

$$T_3=\frac{x-\frac{rc}{N}}{\sqrt{\frac{rc(N-r)(N-c)}{N^2(N-1)}}} \tag{7}$$

得到一个大样本逼近，其渐近分布服从在表A1中给出的标准正态分布．如果行总

和或列总和，或两者同时是随机的，在大样本渐近中用由（1）式给出的 T_1 更精确. 188

假设　令 p_1 为第一行中的一个观测归入第一列的概率，p_2 为第二行中的一个观测归入第一列的概率，t_{obs}为 T_2 的观测值.

A.（双边检验）

$$H_0: p_1 = p_2$$
$$H_1: p_1 \neq p_2$$

首先用（6）式来求 p-值. p-值是 P（$T_2 \leqslant t_{obs}$），P（$T \geqslant t_{obs}$）中较小者的 2 倍. 在显著性水平 α 下，如果 p-值$\leqslant\alpha$，则拒绝 H_0.

B.（左边检验）

$$H_0: p_1 \geqslant p_2$$
$$H_1: p_1 < p_2$$

用（6）式求出 p-值 P（$T_2 \leqslant t_{obs}$）. 在显著性水平 α 下，如果 p-值$\leqslant\alpha$，则拒绝 H_0.

C.（右边检验）

$$H_0: p_1 \leqslant p_2$$
$$H_1: p_1 > p_2$$

用（6）式找到 p-值 P（$T_2 \geqslant t_{obs}$）. 在显著性水平 α 下，如果 p-值$\leqslant\alpha$，则拒绝 H_0.

计算机辅助　Fisher 精确检验可以在 *S-Plus*，*SAS*，*StatXact* 中找到. ◄

评注

这个检验对于具有随机行总和，随机列总和，或行列总和同时都是随机的列联表是有效的. 也就是说，这个精确检验为样本空间中给定行总和与列总和的一个子集求出了 p-值. 每个不同的行列总和集又表示另外的互不相容的子集，因此将整个样本空间分成了几个互不相容的子集. 如果每个子集的临界域在假设 H_0 下，有一个条件概率小于等于 α，那么所有临界域的并在 H_0 下有一个无条件概率小于等于 α，并且检验是有效的. 然而，这种精确检验的功效通常小于一种更适当的近似行总和，或列总和，或行列总和为随机的检验功效. 189

评注（连续性修正）

T_3 的大样本逼近可以通过连续性修正来改进. 即对于左边概率，在从表 A1 中查 p-值之前，给 T_3 的分子加上 0.5，对于右边概率，则从分子上减去 0.5. 这样得到的概率在多数情况下会更精确.

例 4.1.3

银行新雇用了 10 男 4 女共 14 个员工，能力等同. 银行主管正在给他们分配新工作，有 10 个岗位是出纳员，4 个是账户代表. 零假设是男女有等同的机会得到想要的账户代表的工作，单边备择假设是女性比男性更有可能得到账户代表的工作.

只有一位女性被分配为出纳员，那么零假设可以拒绝吗？因为行列总和已确定，是非随机的，所以可以填入下面的 2×2 列联表.

	账户代表	出纳员	
男性	1	9	10
女性	3	1	4
	4	10	14

$$H_0: p_1 \geqslant p_2$$
$$H_1: p_1 < p_2$$

由（6）式给出的精确的左边 p-值是：

$$P(T_2 \leqslant 1) = P(T_2 = 0) + P(T_2 = 1)$$
$$= \frac{\binom{10}{0}\binom{4}{4}}{\binom{14}{4}} + \frac{\binom{10}{1}\binom{4}{3}}{\binom{14}{4}}$$
$$= \frac{1}{1001} + \frac{40}{1001} = 0.041$$

190 当 $\alpha = 0.05$ 时，拒绝零假设. ■

评注

将例4.1.3中的精确 p-值与列总和是随机情况下的精确 p-值作一下比较，即假定一个问题中它的列总和是随机的、并且结果与上面所给出的列联表相同，试验者想通过（4）式得到精确的 p-值. 将 $p = 0.03$ 代入（4）式（见思考题3），可以得到 T_1 的左边概率的最大值，其精确 p-值是0.012，它比之前用 Fisher 精确检验得到的0.041小的多. 从该例中列联表得到 $T_1 = -2.4321$，查表A1得到 T_1 的正态渐近为0.008，这接近于 p-值的真值. 这说明 Fisher 精确检验是精确的仅当行列总和是非随机的. 在其他情形下，Fisher 精确检验仍有效，但有引起极大争议的趋势.

□**理论** 为了说明 T_2 服从超几何分布，让我们从一个具有固定行总和的列联表开始，其概率由（4）式给出如下（变换记号）：

$$\binom{r}{x}\binom{N-r}{c-x} p^c (1-p)^{N-c} \tag{8}$$

取值为 c，$N-c$ 的列总和服从二项分布，其概率如下：

$$\binom{N}{c} p^c (1-p)^{N-c} \tag{9}$$

同例1.3.8一样，在给定列总和的条件下，得到表上结果的条件概率，可用（8）式除以（9）式得到（6）式得到. 通过在 T_2 上减去均值除以超几何分布的标准差得到 T_3，并用中心极限定理得到大样本正态逼近. □

有时需要将几个 2×2 列联表合成一个做整体分析. 当一个整体试验包括几个在不同环境中操作的小试验时，在零假设下的共同的概率随着环境的不同而不同，

并且每一个小试验都有自己的 2×2 列联表，这时常常需要进行这种处理．因为得到每个列联表的环境不同，所以这几个表不能合成单一的一个 2×2 列联表．

Mantel 和 Haenszel(1959)提出了合并几个 2×2 列联表的一种方法． 191

▶Mantel-Haenszel 检验

数据　将数据综合以后放入几个 2×2 列联表中，每个列联表的行列总和都是非随机的．

假设表的数目 $k \geq 2$，并且第 i 个表具有如下形式：

	列 1	列 2	
行 1	x_i	$r_i - x_i$	r_i
行 2	$c_i - x_i$	$N_i - r_i - c_i + x_i$	$N_i - r_i$
	c_i	$N_i - c_i$	N_i

假定条件　每个列联表的假设条件与 Fisher 精确检验相同，并且几个列联表是由独立的试验得到的．

检验统计量

$$T_4 = \frac{\sum x_i - \sum \frac{r_i c_i}{N_i}}{\sqrt{\sum \frac{r_i c_i (N_i - r_i)(N_i - c_i)}{N_i^2 (N_i - 1)}}}$$

零分布　若零假设为真，T_4 的分布近似于表 A1 中给出的标准正态分布，并且可以通过连续修正来提高精确的概率．也就是说，对于左边概率，在从表 A1 中查 p-值时给 T_4 的分子加上 0.5，对于右边概率，减去 0.5，这样得到的概率在多数情况下会更精确．

假设　在第 i 个列联表中，令 p_{1i}是被归入第一行第一列中的观测的概率，令 p_{2i}是第二行第一列相应的概率．

A．（双边检验）

H_0: $p_{1i} = p_{2i}$，对所有的 $i = 1, 2, \cdots, k$

H_1: 或 $p_{1i} > p_{2i}$，对某个 i，或 $p_{1i} < p_{2i}$ 对某个 i，但两个不同时成立． 192

在水平 α 下，如果 T_4 大于 $z_{1-\alpha/2}$，或 T_4 小于 $z_{\alpha/2}$ 则拒绝零假设．这里 z_p 代表的是表 A1 给出的标准正态分布的 p 分位数．

一个服从标准正态分布的随机变量小于观测 T_4 或大于观测 T_4 的概率，不论哪一个小，p-值都 2 倍于此概率．

B．（左边检验）

H_0: $p_{1i} \geq p_{2i}$，对所有的 $i = 1, 2, \cdots, k$

H_1: $p_{1i} \leq p_{2i}$，对所有的 i，且对某个 i，$p_{1i} < p_{2i}$

在水平 α 下如果 T_4 小于 z_α，则拒绝零假设，z_α 从表 A1 中获得．p-值是一个服从标

准正态分布的随机变量小于观测 T_4 的概率.

C. （右边检验）

$$H_0: p_{1i} \leqslant p_{2i}, \text{对所有的} i = 1, 2, \cdots, k$$
$$H_1: p_{1i} \geqslant p_{2i}, \text{对所有的} i, \text{且对某个} i, p_{1i} > p_{2i}$$

在水平 α 下如果 T_4 大于 $z_{1-\alpha}$，则拒绝零假设，$z_{1-\alpha}$可从表 A1 中获得. p-值是一个服从标准正态分布的随机变量大于观测 T_4 的概率.

计算机辅助　Mantel-Haenszel 检验可以在 *S-Plus* 和 *SAS* 中找到. ◄

评注

像 Fisher 精确检验一样，即使行总和或列总和是随机的，这种检验仍是有效的. 但在那种情况下，用如下的检验统计量代替 T_4 更准确.

$$T_5 = \frac{\sum x_i - \sum \frac{r_i c_i}{N_i}}{\sqrt{\sum \frac{r_i c_i (N_i - r_i)(N_i - c_i)}{N_i^3}}}$$

它可以与正态分布作比较，正如像上面描述的 T_4 一样，当用检验统计量 T_5 时，不应

193 该用连续修正来找 p-值.

例 4.1.4

参考 Li，Simon 和 Gart(1979)，在一个测试成功率是否提高的试验性治疗中，把癌症病人分成 3 组. 成功与失败的次数概括如下：

	组 1		组 2		组 3	
	成功	失败	成功	失败	成功	失败
治疗	10	1	9	0	8	0
控制	12	1	11	1	7	3

因为 p_{1i}代表了治疗第 i 组病人成功的概率，所以用右边检验：

$$T_4 = \frac{(10+9+8) - \left(\frac{11 \cdot 22}{24} + \frac{9 \cdot 20}{21} + \frac{8 \cdot 15}{18}\right)}{\sqrt{\frac{11 \cdot 22 \cdot 13 \cdot 2}{24^2 \cdot 23} + \frac{9 \cdot 20 \cdot 12 \cdot 1}{21^2 \cdot 20} + \frac{8 \cdot 15 \cdot 10 \cdot 3}{18^2 \cdot 17}}}$$
$$= \frac{1.6786}{1.1719} = 1.4323$$

从表 A1 中看到，T_4 没有超过 0.95 分位数 1.6449，所以接受零假设. 把 T_4 的分子减去 0.5，得到 T_4 的连续修正是 1.0057，从而可找到 p-值，从表 A1 中得到右边 p-值是 0.157.

显然在本例中列总和是随机的，用 $T_5 = 1.4690$ 可以得到一个更精确的检验，得到的右边 p-值是 0.071. 这说明了在随机行总和或随机列总和，或行列总和同时随机的情况下，用 T_5 更恰当. 然而，零假设在 $\alpha = 0.05$ 时仍被接受. ■

□**理论** 对每一个列联表的第一行，第一列求和可以得到 T_4 的分子值，也就是 Fisher 精确检验统计量 T_2，然后像 T_3 那样，减去它的均值，除以它的标准差（方差开方）. 由中心极限定理可得，T_4 渐近于标准正态分布.

按照 T_3 的形式，重新排列 T_1，可得到统计量 T_5，T_1 与 T_3 的唯一不同是分母用 N 代替了 $N-1$. 那么仿照前面的推导过程，可以证明 T_5 的正态逼近.

典型的连续修正是取离散随机变量相邻值之间距离的一半，因为行列总和没有改变，所以 T_3 和 T_4 的分子按照一个单位分隔取值时，分母仍为常数. 因此分子中半个单位的连续修正是合适的. 194

但是，若行总和或列总和是随机的，T_1 和 T_5 的分母就会有很多不同的值. 同样地，T_1 和 T_5 也就有很多不同的值，这些值不是平均间隔的. 一个连续修正几乎不可能计算，并且比 0.5 小得多，这种情况下可不做任何修正. 当然，通常所推荐的 0.5 的修正太大了，多数情形下会给出真实概率的相当糟糕的估计. 参考 Pearson(1947)，Plackett(1964)，Grizzle(1967) 和 Conover(1974) 对这一结论的支持. 关于 Mantel-Haenszel 检验的更多信息可以参考 Li，Simon 和 Gart(1997) 及 Breslow 和 Liang(1982). □

遵循 3.1 节中描述的步骤，任何同 2×2 列联表或者任何与列联表相关的未知概率的置信区间都可以得到. 同样，只要假定合理，检验的假设恰当，3.1 中的检验就可以用于列联表.

单边检验的一个简洁法则可以参考 Ott 和 Free(1969). 关于连续修正的进一步探讨可以参考 Mantel 和 Greenhouse(1968)，Pirie 和 Hamdan(1972)，Maxwell(1976)，检验功效的讨论见 Harkness 和 Katz(1964)，精确检验的探讨可参考 Gail 和 Gart(1973)，Garside 和 Mack(1976)及 Madonald，Davis，Milliken(1977). 合成几个 2×2 列联表检验统计量方法参见 Radhakrishna(1965)，Nelson(1966)，Meeker(1978)，和 Zelen(1971). 很多文章讨论了由于误分带来的边际总和的可能误差，其中有 Chiacchierini 和 Arnold(1977)及 Plackett(1977). 其他相关的论文可参考 Fienberg 和 Gilbert(1970)，Upton 和 Lee(1976)及 Ray(1976). Fleiss(1973)的一本优秀的著作里主要讨论了2×2列联表的情形.

习题

1. 为了评估公众对决议立法的反应，从两个总体中各抽取含有 135 人的随机样本. 第一个样本中有 43 人“反对”；第二个样本中有 37 人“反对”. 是不是两个总体中“反对”的人数所占比例不同？与 3. 5 节习题 1，2 作比较，在可能的情况下，可提出两个样本中用相同的人的好处吗？如“前”和“后”两种情况.
2. 60 个学生被平均分到两个班级（每班 30 人）中学习如何编写计算机程序. 一个班级采用传统的学习方法，另一个班级采用试验性的新方法. 在课程结束后，每个学生都参与编程测试，程序要么正确，要么错误，结果列表如下： 195

	正确程序	错误程序
传统班	23	7
试验班	27	3

有理由相信试验性方法优于传统方法吗？或者前面的差异可能是因为偶然的波动吗？

3. 100个男人和100个女人参与试用新牙膏，并说出他们是否喜欢这种牙膏. 32个男人和26个女人说他们不喜欢这种牙膏. 这是否说明大体上男人与女人的偏好有差异？

4. 列联表可以用来描绘度量尺度高于名义尺度的数据. 例如，一个20个观测样本随机抽自美国公民的毕业生，他们的平均等级分数如下：

3.42	3.54	3.21	3.63	3.22	3.80	3.70	3.20	3.75	3.31
3.86	4.00	2.86	2.92	3.59	2.91	3.77	2.70	3.06	3.30

并且，另外一个20个观测样本随机抽自非美国公民的毕业生，他们的平均等级分数如下：

3.50	4.00	3.43	3.85	3.84	3.21	3.58	3.94	3.48	3.76
3.87	2.93	4.00	3.37	3.72	4.00	3.06	3.92	3.72	3.91

检验零假设：美国公民的毕业生平均等级分数为3.50或更高的毕业生比例与非美国公民毕业生的比例一样.

5. Fisher的精确检验可以快速地检验两个变量 X, Y 的相关性，每一个变量至少有一个度量的须序尺度. 在 X 的中位数处垂直作一条直线，在 Y 的中位数处平行作一条直线，将 N 个 (X,Y) 值的散点图分成4部分，然后计算每一块中所含点的个数. 注意行总和与列总和是 $N/2$，所以不是随机的.

假设 X 为丈夫结婚的年龄，Y 为他的父亲结婚的年龄，共有16对观测值，两结婚年龄都在中位数以上的有7对，这两个变量是否正相关？

6. 仿习题5，对例3.5.4中的数据用Fisher精确检验来检验正相关性. 记录10个病人对药物1 (X) 和药物2 (Y) 的反应数据如下：(0.7,1.9)，(−1.6,0.8)，(−0.2,1.1)，(−1.2,0.1)，(−0.1，−0.1)，(3.4,4.4)，(3.7,5.5)，(0.8,1.6)，(0.0,4.6)，(2.0,3.4)

比较用Fisher精确检验得到的 p-值和用3.5节中作为相关性检验的Cox和Stuart检验得到的 p-值.

196 7. 怀疑工作中长期接触麻醉剂氧化氮是使怀孕护士和助理牙医流产的原因. 以下数据来自3组不同的怀孕女性，它记录了流产的和足月分娩的人数.

	牙医助理		手术室护士		门诊护士	
	流产	足月	流产	足月	流产	足月
曝光	8	32	3	18	0	7
非曝光	26	210	3	21	10	75

(a) 用 T_4 来检验这个论断，求 p-值时用连续性修正.

(b) 用 T_5 来检验流产不是受氧化氮影响的假设，比较(a)，(b)的 p-值.

(c) 这个案例中用 T_4 或 T_5，哪个分析更合理？

8. (a) 一所大学去年接到21位男性和63位女性的求职信，结果聘用了10位男性与14位女

性，比例分别为 48% 和 22%．这所大学聘用男性的概率是否比聘用女性的概率更大（用 Fisher 精确检验）？

（b）根据学院详细分类的数据如下：

申请者	教育学院 被聘	教育学院 被拒	管理学院 被聘	管理学院 被拒	工程学院 被聘	工程学院 被拒
男性	2	8	5	0	3	3
女性	12	48	1	0	1	1

这所大学聘用男性的概率是否比聘用女性的概率更大（用 Mantel – Haenszel 检验）？

（c）（a），（b）中结论有差异的原因是什么？试讨论之．

思考题

1. 在概率差异检验中，当 $n_1=2, n_2=3$ 时，求出检验统计量的精确概率分布．令 T_1 的最大值对应临界域，并求出 α.

2. 本节的数据可以看成从总体 1 和总体 2 中抽取的两个独立样本，分别包含观测 $X_1, X_2, \cdots, X_{n_1}$ 和 $Y_1, Y_2, \cdots Y_{n_2}$. 如果观测在类 1 中，则每个 X_i 或 Y_i 为 0，如果在类 2 中，则为 1. 因此每个样本就是一组 0 和 1. 对两个独立样本问题的参数方法就是用"两样本 t 检验"，其统计量为： 197

$$t=\frac{\overline{X}-\overline{Y}}{\sqrt{n_1S_x^2+n_2S_y^2}}\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}$$

这里 $\overline{X}, \overline{Y}$ 是两样本的样本均值，S_x^2, S_y^2 是两样本的样本方差. 证明 t 和 T_1（(1)式）的关系如下：

$$t=T_1\sqrt{\frac{n_1+n_2-2}{n_1+n_2-T_1^2}}$$

或等价为

$$T_1=t\sqrt{\frac{n_1+n_2}{n_1+n_2-2+t^2}}$$

这种关系表明，如果 t 和 T_1 的临界域一致，对于较大的 t 拒绝 H_0，等价于对较大的 T_1 拒绝 H_0.

3. 考虑 $n_1=10$ 和 $n_2=4$ 的两个随机样本，检验问题 $H_0: p_1=p_2, H_1: p_1 \leqslant p_2$ 为左边检验. 观测值的 2×2 列联表如下：

	类 1	类 2	
总体 1	1	9	10
总体 2	3	1	4
	4	10	14

（a）计算检验统计量 T_1.

（b）对 $n_1=10$ 和 $n_2=4$，找 4 个列联表使 T_1 值更小（更负）．还能找到多于 4 个列联表吗？

（c）用（4）式，求出精确的 p-值，即：（b）中观测数据表的概率和多于 4 个表的概率，

它是假设 H_0 下共同概率 $p=p_1=p_2$ 的函数.

(d) 将 $p=0.3$ 代入 (c) 中所得的概率等式中，求左边检验中与观测表相关的 p-值.（提示：应得到 $p=0.012$，如本节所做的那样）

(e) 分别将 $p<0.3$ 和 $p>0.30$ 的一个值代入 (c) 中所得的概率等式中，说明在 $p=0.3$

198 处 p-值最大.

4. 证明 T_1，T_3 具有如下关系式：

$$T_3=\frac{\sqrt{N-1}}{\sqrt{N}}T_1$$

4.2 $r\times c$ 列联表

将上节的 2×2 列联表直接推广到具有 r 行 c 列的列联表，即 $r\times c$ 列联表. 同上节一样，$r\times c$ 列联表可以用来描绘包含几个样本的数据，这里的数据至少表示一个度量的名义尺度，也可以用来检验概率不随样本改变而改变的假设. $r\times c$ 列联表的另一个用途是用于单个样本，这个样本中的每个元素根据一种标准可以归入 r 个不同类之一，同时也可以根据另一个标准归入 c 个不同类之一. 在统计分析中视这两种用途一样，但是由于它们存在根本的差异，所以需要将两种情形分别讨论. 本节也将讨论第三种应用，它类似于前两种.

首先考虑将上节介绍的两样本的应用扩展. 现在我们把两个样本推广成 r 个样本，每个样本为 r 行中的一行，每个样本的每个观测可以由前面的两类（类1，类2）推广到归入 c 类，对应着 c 列. 于是可将第 i 个样本的第 j 个值填入 (i,j) 格（第 i 行，第 j 列）中. 因为行列数目多，所以上节的单边假设不再适用，因此从上面推广到 $r\times c$ 情形时，我们仅仅考虑双边备择假设及检验统计量是 T_1 的平方.

▶概率差异的 χ^2 检验，$r\times c$ 情形

数据 共有 r 个总体，从每个总体中抽取一个随机变量. 记第 i 个样本含有的观测数为 n_i，$1\leqslant i\leqslant r$. 每个样本的每个观测可以归入 c 个不同类中的一类. 记 O_{ij} 为样本 i 的观测归入类 j 的数目，所以，

$$n_i=O_{i1}+O_{i2}+\cdots+O_{ic}\quad 对所有 i \tag{1}$$

199 将数据排列成以下 $r\times c$ 列联表中.

	类 1	类 2	···	类 c	总和
总体 1	O_{11}	O_{12}	···	O_{1c}	n_1
总体 2	O_{21}	O_{22}	···	O_{2c}	n_2
···	···	···	···	···	···
总体 r	O_{r1}	O_{r2}	···	O_{rc}	n_r
总和	C_1	C_2	···	C_c	N

所有样本的观测总数记为 N.

$$N = n_1 + n_2 + \cdots + n_r \tag{2}$$

第 j 列中的观测数记为 C_j，即 C_j 是所有样本在第 j 列中的观测总数.

$$C_j = O_{1j} + O_{2j} + \cdots + O_{rj}, \quad 对\ j = 1, 2, \ldots, c \tag{3}$$

假定条件

1. 每个样本都是一个随机样本.

2. 不同样本的输出结果是相互独立的（特别对样本之间，样本之内由假设条件 1 知是独立的）.

3. 每个观测只能归入 c 类中的一类中.

检验统计量 给定检验统计量 T 为：

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \qquad 其中 \qquad E_{ij} = \frac{n_i C_j}{N} \tag{4}$$

这时如果 H_0 为真，O_{ij}代表格（i,j）的观测数，E_{ij}代表格（i,j）期望的观测数，即如果 H_0 为真，格（i,j）的观测数应接近第 i 个样本大小 n_i 与所有观测归入 j 类的比例 C_j/N 的乘积. 注意在 2×2 情形下，因为仅仅考虑双边备择假设，这里的 T 等于上节的 T_1^2.

为了方便计算，我们给出 T 的一个等价的表达式：

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2}{E_{ij}} - N \tag{5}$$ 200

零分布 T 的零分布是渐近自由度为（$r-1$）（$c-1$）的χ^2 分布. χ^2 分布值见表 A2. T 的精确分布很难算出，所以几乎不用.

如果 E_{ij}在检验统计量中不是很小，χ^2 近似是很好的. 一般来说，如果所有的 E_{ij}大于 0.5 且至少一半的 E_{ij}大于 1.0，则χ^2 近似效果较为满意，但若存在一个 E_{ij}小于 0.5 或大多数的 E_{ij}小于 1.0 时，χ^2 近似就未必精确了. 但此时我们可以将行或列加和起来以便于去除那些值很小的情况；或者去除有很小观测值的行或列.

假设 记 p_{ij}为随机取到第 i 个总体划分为第 j 个类的概率，$i=1,2,\cdots,r, j=1,2,\cdots c$.

H_0: 同一列中的所有概率相等（对任意j, $p_{1j} = p_{2j} = \cdots = p_{rj}$）.

H_1: 每列中至少存在两个概率相等（即对于给的j存在一对i与k，使得$p_{ij} \neq p_{kj}$）.

注意，没有必要来规定各种概率，零假设只表述对所有总体归入第 j 类的概率是相同的，无论这概率是什么，也不管我们考虑的是哪一个类型.

因为计算 T 的精确分布很困难，我们用大样本（这里 E_{ij}较大）分布逼近去找临界域. 在近似水平 α 下的临界域对应于 T 值大于 $X_{1-\alpha}$，这里 $X_{1-\alpha}$是自由度为($r-1$)($c-1$)的χ^2 分布的 $1-\alpha$ 分位数，它可由表 A2 获得. 当 T 大于 $X_{1-\alpha}$时，拒绝 H_0，否则接受 H_0.

p-值是自由度为($r-1$)($c-1$)的χ^2 随机变量大于 T 的概率，可查表 A2 获得.

计算机辅助 *Minitab*，*S-plus*，*SAS*，*Statxact* 都可以做这个检验，有些程序可分析比本书更为复杂的列联表. ◀

评注（E_{ij}较小的情形）

因为使用了渐近分布，如果 E_{ij}适当大，α 的近似值就很接近 α 的真值，然而，
201 如果某些 E_{ij}较小时，这种近似效果可能非常差.

Cochran(1952)发现，如果存在 E_{ij}小于 1 或者超过 20% 的 E_{ij}小于 5，那么这种近似可能很差. 但根据很多学者未发表的研究表明，这近乎太保守了. 其中包括 B. L. Vander Waerden 的学生和 Oscar Kempthorne 的学生的一些研究，以及 Roscoe 和 Byars(1971)年的文章. 如果 r 与 c 不太小的话，我认为即使一些 E_{ij}小于 0.5，但大部分的 E_{ij}大于 1，检验也可以是有效的. 如果一些 E_{ij}值太小，可合并几个类去消除 E_{ij}太小的影响，需要决定的是哪一些类需要合并. 一般来说，如果某些类在某些方面很相似，我们就可合并它们而保留了原假设的意义.

例 4.2.1

随机地从私立中学与公立中学里抽取一些学生进行标准测验，得到如下结果.

测验分数

	0–275	276–350	351–425	426–500	总和
私立中学	6	14	17	9	46
公立中学	30	32	17	3	82
总和	36	46	34	12	128

为检验零假设：私立中学与公立中学的学生的分数服从相同的分布，我们可用概率的差异性检验. 近似水平 $\alpha = 0.05$ 下的拒绝域对应于 T 大于 7.815 的值，这从 A2 表得到，其中 χ^2 分布的自由度是 $(r-1)(c-1)=(2-1)(4-1)=3$.

用(4)式计算 E_{ij}的值如下：

列

E_{ij}:	1	2	3	4
行 1	12.9	16.5	12.2	4.3
行 2	23.1	29.5	21.8	7.7

注意，这里 E_{ij}满足 Cochran 的条件，也值得注意的是 E_{ij}行列的和总是与 O_{ij}行列的
202 和相同. 这对检查计算是有用的.

对于第一行，第一列所在的格子，我们有：

$$\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \frac{(O_{11}-E_{11})^2}{E_{11}} = \frac{(6-12.9)^2}{12.9} = \frac{47.61}{12.9} = 3.69$$

对其他格子可进行同样的计算，结果如下：

$$T = 3.69 + 0.38 + 1.89 + 5.14 + 2.06 + 0.21 + 1.06 + 2.87 = 17.3$$

因为 $T=17.3$ 大于7.815. 所以我们拒绝零假设. 事实上，在0.001这么小的显著性水平下，我们也可以拒绝零假设，故 p-值大致为0.001.

结果是私立中学与公立中学的测试分数的分布是不同的. ■

在例4.2.1中的数据（分组前的测试分数）至少是度量的顺序尺度. 一个强于度量名义尺度的尺度对测试分数所用的检验可能更合适. 如果感兴趣的备择假设是私立中学的学生比公立中学的学生可获得较高（低）的分数，则我们可用基于秩的更有效的检验. 比如下一章中介绍的Mann-Whitney检验. 然而，本例中的备择假设包含了所有类型的差异，比如高分，低分，分数内的小变异，分数内的大变异等等，所以这里用 χ^2 检验更恰当.

□**理论**　仿照上节 2×2 情形，可以找到 $r\times c$ 情形中 T 的精确分布，即令行总和（样本大小）保持常数，然后列举所有可能的具有相同行总和的列联表，对每一行用多项分布计算它们的概率. 列总和可以随表的不同而不同，但是行总和不能变，这一点是和下面将要描述列联表应用的本质区别. 在以下的应用中，没有固定行总和，因此，有很多可能的列联表，仅要求对所有表都有一样的观测总数 N. 并且提一下第3种变化，在这种应用中，行列总和都不随表的变化而变化，是固定的，因此可能列联表的数量被大大缩减，精确的分布也更容易求得.

在本节列联表的3种应用中，T 的渐近分布是一样的，都是自由度为 $(r-1)(c-1)$ 的 χ^2 分布，因此可用这个分布来给 α 一个近似值，而不必需要知道精确的列 203
表值. Cramér(1946)推导出了渐近分布. □

对于含有容量为 N 的单个随机变量，每个观测可以根据两个准则来分类的情形，可以用 $r\times c$ 列联表，这也是 $r\times c$ 列联表的第二种应用. 由第一个准则产生 r 类（行），由第二个准则产生 c 类（列）. 将每个观测按照两条准则分类后，填入 $r\times c$ 列联表的一个对应的格子中去. 归入的格表示属于该格子的观测数，尽管会用较高的度量尺度，但一个度量至少需要是名义尺度. 假设检验是一种独立性的检验；繁赘一点地说，零假设为行与列代表了两个独立的分类方案. 下面我们给出更精确的描述.

▶独立性的 χ^2 检验

数据　已知一个容量为 N 的随机样本，它的观测值根据两个准则划分成几类. 按照第一个准则，每个观测可归入 r 类（行）中的某一类，按照第二个准则每个观测可归入 c 类（列）中的某一类. 记 O_{ij} 为第 i 行第 j 列的观测数，将其填入 $r\times c$ 列联表中. 记第 i 行的观测总数为 R_i，（代替前面检验中的 n_i，是为了强调行总和是随机的而不是固定的），记第 j 列观测总数为 C_j，所有格中总的观测数为 N.

	列 1	2	3	···	c	总和
行 1	O_{11}	O_{12}	O_{13}	···	O_{1c}	R_1
2	O_{21}	O_{22}	O_{23}	···	O_{2c}	R_2
···	···	···	···	···	···	···
r	O_{r1}	O_{r2}	O_{r3}	···	O_{rc}	R_r
总和	C_1	C_2	C_3	···	C_c	N

假定条件

1. 观测为 N 的样本是随机的.（每个观测归入第 i 行第 j 列的概率相同，并且同
204 其他观测独立）

2. 按照第一个准则，每个观测可归入 r 类（行）中的某一类，按照第二个准则，每个观测可被归入 c 类（列）中的某一类.

检验统计量 令：$E_{ij}=R_iC_j/N$，则给出检验统计量如下

$$T=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \tag{6}$$

或者为了手算的方便，

$$T=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2}{E_{ij}}-N \tag{7}$$

这里的求和是对列联表的所有格子求和. 注意，(6) 和 (7) 式与前面 (4) 和 (5) 式两种检验形式相同.

零分布 正如前面的检验，T 的零分布可以通过自由度为 $(r-1)(c-1)$ 的 χ^2 分布（可以在表 A2 中查到）来近似. T 的精确分布难以求得，因此几乎不用.

正如前面的检验，如果所有的 E_{ij} 都大于 0.5，并且至少有一半 E_{ij} 大于 1.0，那么 χ^2 近似通常是令人满意的. 将总和较小的行或列与其他描绘相似特征的行或列合并，或者简单删除几乎不包含观测的行或列，可以消除小的 E_{ij} 的影响.

假设

H_0：对任意的 i，j，事件"一个观测值在行 i"与事件"同样的观测在列 j"是独立的.

通过事件独立性的定义，H_0 可以表述如下：

$$H_0: P(\text{行} i, \text{列} j)=P(\text{行} i)\cdot P(\text{列} j), \text{对所有的} i, j$$

备择假设可以方便地表述为：

$$H_1: P(\text{行} i, \text{列} j)\neq P(\text{行} i)\cdot P(\text{列} j), \text{对某} i, j$$

查表 A2，如果 T 超过自由度为 $(r-1)(c-1)$ 的 χ^2 分布随机变量的 $1-\alpha$ 分位数，则拒绝 H_0，置信水平为 α. p-值（自由度为 $(r-1)(c-1)$ 的一个服从 χ^2 分布的随机变

量超过观测值 T 的概率）也可以从表 A2 中查到.

计算机辅助 本检验可用 *Minitab*, *S-Plus*, *SAS* 和 *StatXact* 操作完成，这些软件里的一些程序还可以分析比本书中更复杂的列联表. 205

例 4.2.2

将某个大学里学生作为一个样本，根据他们被录取的院系和是否从州内或州外的高中毕业两个标准来分类. 将结果填入到如下 2×4 列联表中，

	工程学院	艺术与科学学院	国内经济学院	其他	总和
州内	16	14	13	13	56
州外	14	6	10	8	38
总和	30	20	23	21	94

为了检验零假设，每个学生被录取的院系与他们是否在州内还是州外读高中独立，我们选择 χ^2 独立性检验来检验这个假设. $N=4$ 查表 A2，自由度为 $(r-1)(c-1)=3$ 的 χ^2 分布随机变量的 0.95 分位数为 7.815，故近似水平为 0.05 的拒绝域是 $(7.815,\infty)$，因此 α 近似于 0.05.

用（6）式算出 T 为：

$$T=1.52$$

因此可以接受零假设. 从表 A2 中查到 p-值大于 0.25. ■

□**理论** 我们可以给出相对简单情形 $N=4$ 时的 T 的精确分布. 令 p_{ij} 是一个观测归入第 i 行第 j 列的概率.（注意这里的 p_{ij} 与前面检验中的 p_{ij} 不同，所有格子中的 p_{ij} 之和是 1，而前面检验中的 p_{ij} 每一行加起来为 1）. 则有如下特殊结果

	列 1	列 2	
行 1	a	b	
行 2	c	d	
			N

206

的概率可以用多项式分布求得：

$$\frac{N!}{a!b!c!d!}(p_{11})^a(p_{12})^b(p_{21})^c(p_{22})^d \tag{8}$$

因为 N 个对象能导出以上列联表的个数由多项式系数 $\frac{N!}{a!\,b!\,c!\,d!}$ 给出，并且每个结果的概率为：

$$(p_{11})^a(p_{12})^b(p_{21})^c(p_{22})^d \tag{9}$$

零假设：每个 p_{ij} 等于它的行概率乘上列概率，当 H_0 为真时，由所有的 p_{ij} 相等（本情形中 $p_{ij}=1/4$，这个结论我们不证明）可得 T 的右边的最大概率. 因此，对每种可能

的排列，通过计算

$$P\left(\begin{array}{|c|c|}\hline a & b \\ \hline c & d \\ \hline\end{array}\right)=\frac{N!}{a!b!c!d!}\left(\frac{1}{4}\right)^N \tag{10}$$

而得到一个 α，对 $N=4$ 的情形，我们可以找到35种不同的列联表. 这些表及其发生的概率与 T 值都在图4-1中列出. 同前，我们定义 $0/0=0$.

$$P(T=0)=84/256=0.33$$
$$P(T=4/9)=48/256=0.19$$
$$P(T=4/3)=96/256=0.37$$
$$P(T=4)=28/256=0.11$$

如果临界域对应着最大的 T 值，$T=4$，则 $\alpha=0.11$，同上节中讨论的精确分布的双边检验得到的置信性水平 $\alpha=0.125$ 相比，差异不是很大. 同上节中 N 固定，行总和也固定下得到的分布相比，这里 T 的分布更复杂，原因是这里可能的列联表方式更多，并且，T 可能有例外值，因此可能的概率分布也改变了一些. □

尽管两种应用下的 T 的精确分布有少许差异，但是都可由自由度为 $(r-1)(c-1)$ 的 χ^2 分布近似.

列联表的第三种应用中，不仅行总和为固定（同第一种应用一样），而且列总和也固定，因此较前面介绍的两种应用 T 的精确分布更容易求出. 然而，除非有辅助
207 表或计算机的帮助，T 的精确分布对实际应用来说仍太复杂，所以我们建议用 χ^2 近似来求临界值和 α.

▶固定边缘分布的 χ^2 检验

数据　数据归纳入一个 $r\times c$ 列联表中，这与前两种应用一样，不同的是这里的行与列总和固定而非随机.

	列				
	1	2	· · ·	c	总和
行 1	O_{11}	O_{12}	· · ·	O_{1c}	n_1
2	O_{21}	O_{22}	· · ·	O_{2c}	n_2
· · ·	· · ·	· · ·	· · ·	· · ·	· · ·
r	O_{r1}	O_{r2}	· · ·	O_{rc}	n_r
总和	c_1	c_2	· · ·	c_c	N

记行与列总和分别为 n_i 和 c_j，这种记号是为了强调行与列总和是固定而非随机的，观测总数为 N.

$T=0$

结果	概率
4 0 / 0 0	$(1/4)^4$
0 4 / 0 0	$(1/4)^4$
0 0 / 0 4	$(1/4)^4$
0 0 / 4 0	$(1/4)^4$
3 1 / 0 0	$4(1/4)^4$
0 3 / 0 1	$4(1/4)^4$
0 0 / 1 3	$4(1/4)^4$
1 0 / 3 0	$4(1/4)^4$
1 3 / 0 0	$4(1/4)^4$
0 1 / 0 3	$4(1/4)^4$
0 0 / 3 1	$4(1/4)^4$
3 0 / 1 0	$4(1/4)^4$
2 2 / 0 0	$6(1/4)^4$
0 2 / 0 2	$6(1/4)^4$
0 0 / 2 2	$6(1/4)^4$
2 0 / 2 0	$6(1/4)^4$
1 1 / 1 1	$24(1/4)^4$

总和 = 84/256

$T=4/9$

结果	概率
2 1 / 1 0	$12(1/4)^4$
1 2 / 0 1	$12(1/4)^4$
0 1 / 1 2	$12(1/4)^4$
1 0 / 2 1	$12(1/4)^4$

总和 = 48/256

$T=4/3$

结果	概率
2 1 / 0 1	$12(1/4)^4$
0 2 / 1 1	$12(1/4)^4$
1 0 / 1 2	$12(1/4)^4$
1 1 / 2 0	$12(1/4)^4$
2 0 / 1 1	$12(1/4)^4$
1 2 / 1 0	$12(1/4)^4$
1 1 / 0 2	$12(1/4)^4$
0 1 / 2 1	$12(1/4)^4$

总和 = 96/256

$T=4$

结果	概率
3 0 / 0 1	$4(1/4)^4$
0 3 / 1 0	$4(1/4)^4$
1 0 / 0 3	$4(1/4)^4$
0 1 / 3 0	$4(1/4)^4$
2 0 / 0 2	$6(1/4)^4$
0 2 / 2 0	$6(1/4)^4$

总和 = 28/256

图 4-1　当所有的 $p_{ij}=1/4$ 时，T 的精确分布 208

假定条件

1. 每个观测只能归入一个格中.
2. 观测是一个随机样本里的观测，每个观测落入（i，j）的概率相同.
3. 行与列总和是固定而非随机的.

检验统计量 令 $E_{ij}=n_i c_j/N$ 为落入 (i,j) 中的期望观测数，同前，给出检验统计量如下：

$$T=\sum_{i=1}^{r}\sum_{j=1}^{c}(O_{ij}-E_{ij})^2/E_{ij} \tag{11}$$

其中求和是对所有 rc 个格子求和. 注意，这个检验统计量形式同前面所用的两个相
209 同，所以为了计算方便，可用（5）式代替.

零分布 同前面的检验一样，T 的零分布可以用自由度为 $(r-1)(c-1)$ 的 χ^2 分布近似. T 的精确分布虽然比前两种检验容易求得，但还是很困难，所以很少用它.

同前面的检验一样，如果所有的 E_{ij} 比0.5大或至少有一半大于1.0，χ^2 近似通常是令人满意的. 可将一些特征相似，总和较小的行或列合并，或简单地消去几乎没有观测值的行或列来消除较小 E_{ij} 的影响.

假设 这里的假设可以取本节前两种应用中的两组假设之一，但行与列总和必须是固定的，或者取适应特定背景的假设. 通常这种假设是前面检验独立性假设的变化，见下面例4.2.3和例4.2.4根据试验所做的特殊修改.

若 $T>\chi^2_{1-\alpha}((r-1)(c-1))$，则拒绝 H_0，近似的置信性水平为 α，p-值也可从表A2中查到，$p=p(X>T)$，$X\sim\chi^2((r-1)(c-1))$.

计算机辅助 可在 *Minitab*, *S-Plus*, *SAS* 和 *StatXact* 中进行这个检验. 对某些情形，可通过 *StatXact* 求出精确的 p-值.

例4.2.3

固定边际总和的 χ^2 检验也可用来检验两个随机变量 X 和 Y 是否独立. 已知有一个含有24个点的散点图，这些点代表了二元随机变量 (X,Y) 独立的观测，可将图4-2构建成列联表. 每个点的横坐标是随机变量 X 的观测值，纵坐标是 Y 的观测值. 假设每对观测 (X,Y) 是相互独立的，我们希望检验如下问题：

H_0: X 和 Y 是相互独立的，H_1: X 和 Y 是不独立

为了形成一个所有 E_{ij} 都相等的列联表，我们注意到3和4都是样本容量为24的因子，因此我们用虚线将图4-2中两点平均分成3行4列，即每行有8个点，每列有6个点（如果一些 E_{ij} 很小，应令它们几乎相等，一种方式是行总和相等且列总
210 和相等），这样得到的列联表如下：

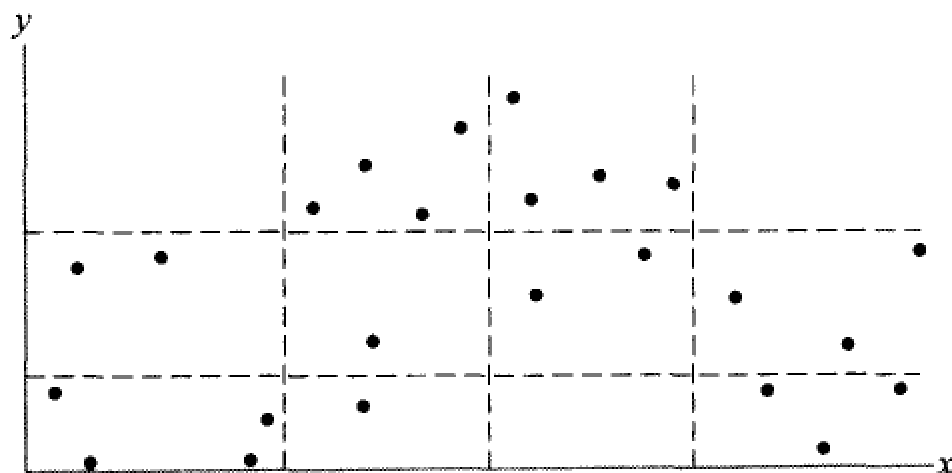


图4-2 散点图

	列 1	2	3	4	总和
行 1	0	4	4	0	8
2	2	1	2	3	8
3	4	1	0	3	8
总和	6	6	6	6	24

近似置信性水平为 0.05 的临界域为 $T > \chi^2_{0.95}((2)(3)) = \chi^2_{0.95}(6) = 12.59$（由表 A2 获得）的区域.

检验统计量可用（11）式和 $E_{ij} = (6)(8)/24 = 2$ 计算.

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \sum_{i=1}^{3}\sum_{j=1}^{4}\frac{(O_{ij}-2)^2}{2} = 14 \tag{12}$$

因为 $T > 12.59$，所以应拒绝 H_0，从而我们可以得到 X 和 Y 不独立. 实际上，可以在一个显著性水平小到 0.03 下拒绝 H_0，所以 p-值为 0.03. ■

例 4.2.4

一位心理学家要求被测人学习 25 个单词，给被测人 25 张蓝色卡片，每一张上有一个单词，其中有 5 个名词，5 个形容词，5 个副词，5 个动词和 5 个介词. 她必须将这些蓝色卡片与 25 张白色卡片配对，这些白色卡片也是每张一词，并且词性与每种词性所包含的单词数同蓝色卡片一样（每个词性，5 个单词）. 允许被测人 5 分钟配卡片（1 张白卡片和每张蓝卡片），5 分钟学习所配对的单词. 然后，她被要求闭上眼睛并且一个一个地给她读白色卡片上的单词，当给她读每个单词时，她尽量提供与所读单词相关的蓝色卡片上的单词. 211

心理学家并不关心她正确回答的单词数，而是关心配对结构以检查它是否表示一种次序. 假设如下：

H_0:没有按照词性配对

H_1:被测人倾向于将蓝色卡片上的某一种词性的单词与白色卡片上某一种词性（不一定相同）的单词配对.

把配对结果归入如下 5×5 列联表：

	名词	形容词	副词	动词	介词	总和
名词		3			2	5
形容词	4	1				5
副词				5		5
动词			5			5
介词	1	1			3	5
总和	5	5	5	5	5	25

选择固定边际总和的χ^2检验. 测试者认为若T较大，表明H_1为真. 边际总和代表了每一类的单词数，这在检验前已经是确定的. 近似水平的临界域为$T>26.30$的区域，其中$26.30=\chi^2_{0.95}((r-1)(c-1))=\chi^2_{0.95}((4)(4))=\chi^2_{0.95}(16)$（由表A2获得），用（11）式可以得到$T$的观测值.

$$E_{ij}=\frac{(5)(5)}{25}=1 \qquad \text{对所有 } i \text{ 和 } j$$

$$T=\sum_{i=1}^{5}\sum_{j=1}^{5}\frac{(O_{ij}-1)^2}{1}=66 \tag{13}$$

212 因为$T=66$，所以拒绝H_0，接受H_1是合理的，p-值小于0.001. ■

□**理论** 在上节2×2情形下，若行与列总和是非随机的，则可以找到T的精确分布为超几何分布，其概率由（4.1.6）式给出，现在我们列出2×2情形下T的精确分布.

若行总和与列总和都等于2，则有如下三种可能的列联表：

表	概率	T
$\begin{array}{\|c\|c\|}\hline 2 & 0 \\ \hline 0 & 2 \\ \hline\end{array}$	$\dfrac{\binom{2}{2}\binom{2}{0}}{\binom{4}{2}}=1/6$	4
$\begin{array}{\|c\|c\|}\hline 1 & 1 \\ \hline 1 & 1 \\ \hline\end{array}$	$\dfrac{\binom{2}{1}\binom{2}{1}}{\binom{4}{2}}=2/3$	0
$\begin{array}{\|c\|c\|}\hline 0 & 2 \\ \hline 2 & 0 \\ \hline\end{array}$	$\dfrac{\binom{2}{0}\binom{2}{2}}{\binom{4}{2}}=1/6$	4

因为T的概率分布是惟一的，所以在这个应用中，H_0是简单假设.

将行总和与列总和固定，可以大大减少可能的列联表数目，从而较前两种应用，T的精确分布更容易找到. 若$r=2$，$c=2$，则检验即为“Fisher精确检验”，可以方便地使用更精确的概率表（Finney,1948），程序化的Fisher精确检验可参考Robertson(1960).

对一般的r，c，固定边际总和的列联表

	列 1	2	···	c	总和
行 1	O_{11}	O_{12}	···	O_{1c}	n_1
2	O_{21}	O_{22}	···	O_{2c}	n_2
···	···	···	···	···	···
r	O_{r1}	O_{r2}	···	O_{rc}	n_r
213 总和	c_1	c_2	···	c_c	N

的精确概率为：

$$概率 = \frac{\begin{bmatrix} n_1 \\ O_{1i} \end{bmatrix}\begin{bmatrix} n_2 \\ O_{2i} \end{bmatrix}\cdots\begin{bmatrix} n_r \\ O_{ri} \end{bmatrix}}{\begin{bmatrix} N \\ c_i \end{bmatrix}} \tag{14}$$

这里的多项式系数同 1.1 节中法则 3 的定义. □

本节与 4.1 节中的列联表称为*双向列联表*（two-way contingency table），原因是观测被分成行与列两个方向. 自然的推广方式为，若观测按照三条或三条以上的准则进行分类，那么数据就可用三向（或多向）列联表的形式进行描述.

为了扩展χ^2 列联表检验，我们将双向检验统计量变换为如下形式：

$$T = \sum_{i,j} \frac{\left[O_{ij} - N\frac{R_i}{N}\frac{C_j}{N}\right]^2}{N\frac{R_i}{N}\frac{C_j}{N}} \tag{15}$$

它具有$(r-1)(c-1)$的自由度. 在一个 r 行，c 列和 t 块的三向列联表中，记块总和为 $B_k, k=1,2,\cdots,t$，行总和为 $R_i, i=1,2,\cdots,r$，列总和为 $C_j, j=1,2,\cdots,c$. 记 N 为观测总数，则：

$$R_i = \sum_{j,k} O_{ijk} \tag{16}$$

$$C_j = \sum_{i,k} O_{ijk} \tag{17}$$

$$B_k = \sum_{i,j} O_{ijk} \tag{18}$$

这里 O_{ijk}代表了第 i 行，第 j 列，第 k 块的观测总数，则期望的观测数记为 E_{ijk}，在行－列－块相互独立的假设为真的情况下，E_{ijk}有如下估计：

$$E_{ijk} = N\frac{R_i}{N}\frac{C_j}{N}\frac{B_k}{N} \tag{19}$$ 214

并可用检验统计量

$$T = \sum_{i,j,k} \frac{(O_{ijk} - E_{ijk})^2}{E_{ijk}} \tag{20}$$

计算，这里的求和是对所有的 $r \cdot c \cdot t$ 个格子求和. 然后用 T 和自由度为 $rct-r-c-t+2$ 的χ^2 分布来做显著性检验，其他多向列联表检验可类似推广.

所谓“对数线性模型”的方法可以成功地用来分析多向列联表，它将在本章的最后一节中讨论. 关于多向列联表的更详细的论述可参考：Goodman(1970), Ireland, Ku 和 Kullback(1969), Ku, Varner 和 Kullback(1971), Koch, Johnson 和 Tolley(1972), Darroch(1974)和 Halperin et al. (1977). Maxwell(1961)写了一本分析列联表的小册子. 列联表估计的重要课题可参考 Fienberg(1970a)，McNeil 和 Tukey(1975)，及 Quade 和 Salama(1975). 若一些数据仅仅被部分地分类，可参考 Chen 和 Fienberg

(1974)或 Hocking 和 Oxspring(1974). 5.2 节中将讨论列联表中一类或同时两类有自然排序的情形，也可参考 Williams 和 Grizzle(1972)，Simon(1974)及 Clayton(1974).

等列总和与固定行总和的 2×3 列联表检验统计量的精确分布可参考 Bennett 和 Nakamura(1963,1964)及 Healy(1969). Ireland 和 Kullback(1968)对给定行与列总和的列联表给出了不同的检验. 列联表的χ^2 检验的功效可参考 Chapman 和 Meng(1966).

Mosteller(1968)关于列联表的文章是一篇相当优秀、可读性极强的调研文章. Haynam 和 Leone(1965)给出了 T 的精确分布的近似. 数据误分可参考 Mote 和 Anderson(1965)合写的论文. 讨论有小频率或零频率格子的列联表可参考 Ku (1963) 和 Sugiura 及 Otake (1968). 关于交互检验的信息可参考 Goodman(1964,1968)和 Bhapkar 及 Koch(1968). 一类双变量列联表型分布可参考 Plackett(1965)，Mardia(1967)和 Steck(1968). 其他检验列联表的方法可参考 Ishii(1960)，Gregory(1961)，Claringbold(1961)，Kullback，Kupperman，Ku(1962)，Diamond(1963)，Mielke 和 Siddiqui(1965)，Hoeffding(1965)，Gart(1966)和 Chacko(1966). 很多关于列联表的文章都举例说明了这种分析的实用性和多样性. 比如可参考 Elston(1970)，Crowley 及 Breslow(1975)，Light 和 Margolin(1971)，Margolin 和 Light(1974)；以及 Shuster 和 Downing(1976). Mantel 和 Haenszel(1959)提出了一种有用的检验 N 个独立的 $2\times r$
215 列联表行－列分布一致性的方法. 其他列联表的应用将在本章其余几节中讨论.

习题

1. 检验下边的观测是否表明所观测的两个变量之间是独立的：(3.6,13)，(4.7,19)，(1.4,9)，(5.5,15)，(4.8,27)，(4.3,14)，(3.0,6)，(4.2,11)，(6.0,24)，(6.8,26)，(4.1,18)，(3.2,9)，(4.0,8)，(1.9,6)，(0.4,7)，(4.9,14)，(5.6,18) 和 (5.6,20). 用这一节的哪个检验?
2. 从 80 场马赛中的每场随机挑选出一匹马，并根据起跑位置和马冲过终点线的位置（第一，第二，等等）对其分类.

		终点			
		1	2	3	其他
起跑位置	1–4	8	6	8	16
	5–9	3	6	5	28

马在赛跑终点的位置依赖于它的起跑位置吗? 用这一节的哪个检验?

3. 在另一个研究中，将三天所有马赛中的所有马根据起跑位置和它们赛跑结束的次序来分类.

		终点			
		1	2	3	其他
起跑位置	1–4	15	14	15	52
	5–9	9	10	9	72

马在赛跑终点的位置依赖于它的起跑位置吗？用这一节的哪个检验？

4. 三位教授讲授统计学导论的大班课. 在学期期末，通过比较成绩来看他们的评分方式是否有显著的差别.

	成绩							
教授	A	B	C	D	F	WP	WF	
Smith	12	45	49	6	13	18	2	
Jones	10	32	43	18	4	12	6	
White	15	19	32	20	6	9	7	216

这些有显著差别吗？你用的是哪种检验？Jones 教授和 White 教授所给的分数有显著不同吗？结果怎么解释？

5. 将美国证券交易所（ASE）股票的一个随机样本和纽约证券交易所（NYSE）股票的随机样本进行比较，看这两个交易的百分率是否有区别. ASE 的 23 种股票中有 11 个 A，11 个 B 和 11 个 C，NYSE 的 35 种股票中有 24 个 A，11 个 B 和没有 C. 你的分析如何？

6. 一个观测组穿过一片丛林区，他们报告了所有真的和假的伪装设备指示. 用了两种类型的伪装，普通的和带图案的. 这个组的报告包括所用的伪装类型和设备的位置，这个组被一位知道指示真假的人监控着，真假指示的结果如下.

	伪装类型	
	带图案的	普通的
错误探测的个数	14	4
正确探测的个数	27	32

被报告的不正确指示的概率有显著不同吗？（注意这个研究并不解决未被发现的设备和被认错伪装类型的指示.）这是什么类型的列联表？

7. 将 30 个毕业生的随机样本依据学院和宗教信仰分类如下，学院和宗教信仰有关系吗？

艺术和科学学院	商学院	工学院	其他学院
天主教	新教	天主教	新教
天主教	天主教	新教	天主教
犹太教	犹太教	新教	天主教
天主教	犹太教	新教	其他
新教	新教	其他	天主教
新教	新教		其他
其他	其他		天主教
新教			犹太教
其他			其他

8. 一电视销售公司用电话追踪对其不同产品的反映，以便判断广告在电视中播出的时间是否与产品销售有关. 反映的个数如下，你的分析如何？这里需要什么样的假设？

	产品				
	钓杆	厨具	音乐CD	健身器	
白天	6	73	55	7	
晚上	14	65	82	8	
周末	21	58	48	8	217

思考题

1. 证明由（4）式和（5）式给出的两种形式的 T 是等价的.
2. 证明如果 $r=2$，$c=2$，那么（4）式等价于（4.1.1）式平方

$$T=\frac{N(O_{11}O_{22}-O_{12}O_{21})}{n_1n_2C_1C_2}$$

3. 在很多计算器中找到用统计量

$$T'=2\sum_{i=1}^{r}\sum_{j=1}^{c}O_{ij}\ln(O_{ij}/E_{ij})$$

代替 T 的一种不同的分析列联表的方法，这里 ln 是自然对数. 这两种检验程序的其他地方完全相同. 在习题3中用 T'，看所得的结果是否与用 T 时的类似.（一般来说，这两种检验不等价，尽管它们在特殊情况下会产生类似的结果.）

4.3 中位数检验

中位数检验是用来验证不同总体中抽取的几个样本是否有同样的中位数. 事实上，本章的中位数检验并不新奇，它只不过是上节所介绍的具有固定行列总和的 χ^2 检验的一种具体应用. 因为它非常有用，所以我们将它单独提出来讨论.

为了检验几个总体（c 个）的中位数是否相同，我们从每个总体中抽取一个样本（度量尺度至少是顺序的，否则“中位数”没有意义.），然后构建一个 $2\times c$ 列联表，使第 i 列有两个元素分别为第 i 个样本中高于和低于总中位数（所有观测的中位数）的两个观测值，然后对此列联表用通常的 χ^2 检验.

▶中位数检验

数据 从 c 个总体中各取一个容量为 $n_i, i=1,2,\cdots,c$ 的随机样本，则可确定联合样本的中位数，即在 $N=n_1+n_2+\cdots+n_c$ 个观测中，恰有一半的观测值超过此数，我们称之为总中位数（grand median）. 令 O_{1i} 为第 i 个样本中超过总中位数的观测数，O_{2i} 为第 i 个样本中小于或等于总中位数的观测数，将频数排列在如下的 $2\times c$ 列联
218 表中：

样本	1	2	···	c	总和
>中位数	O_{11}	O_{12}	···	O_{1c}	a
≤中位数	O_{21}	O_{22}	···	O_{2c}	b
总和	n_1	n_2	···	n_c	N

记 a，b 分别为所有样本中大于总中位数和小于或等于总中位数的观测总数，则有 $a+b=N$，N 为观测总数.

假定条件

1. 每一样本都是随机的.

2. 样本之间相互独立.

3. 度量尺度至少是顺序的.

4. 若所有总体有同样的中位数，则对所有总体而言，一个观测超过总中位数的概率相同，记为 p.

检验统计量 将上节给出的检验统计量变换形式，并注意在行数为 2 的特殊情形下，$O_{2i}=n_i-O_{1i}$，我们可得：

$$T=\frac{N^2}{ab}\sum_{i=1}^{c}\frac{\left(O_{1i}-\frac{n_i a}{N}\right)^2}{n_i} \tag{1}$$

为了计算方便，可采用如下形式：

$$T=\frac{N^2}{ab}\sum_{i=1}^{c}\frac{O_{1i}{}^2}{n_i}-\frac{Na}{b} \tag{2}$$

若 a 近似地等于 b（除非观测中有很多等于总中位数），则可简化检验统计量，化简后的形式如下：

$$T=\sum_{i=1}^{c}\frac{(O_{1i}-O_{2i})^2}{n_i} \tag{3}$$

若 $a=b$，则（3）式给出的 T 的简化形式是精确的，否则就是近似的. 219

零分布 由于 T 的精确分布难以求得，所以常采用大样本逼近来近似 T 的分布（见本节后对 T 的精确分布的理论上的讨论）. 近似的零分布是自由度为 $c-1$ 的 χ^2 分布.

假设

H_0: c 个总体有相同的中位数

H_1:至少有两个总体的中位数不同

近似水平为 α 的临界域是 $T>\chi^2_{1-\alpha}(c-1)$ 的区域，其中 $\chi^2_{1-\alpha}(c-1)$ 为一个服从自由度为 $c-1$ 的 χ^2 分布的随机变量的 $1-\alpha$ 分位数，它可以从表 A2 中查到. 若 $T>\chi^2_{1-\alpha}(c-1)$，则拒绝 H_0，否则接受 H_0.

近似的 p-值是一个服从 χ^2 分布的随机变量大于观测值 T 的概率，也可以从表 A2 中查到.

如果一些样本容量 n_i 太小，则上述逼近可能不是很准确（真实的 α 可能与上面近似的 α 差异很大）. 上节中给出的法则可以作为经验法则来用，这时，若在分析中丢掉所有容量为 1 的样本，则上述规则仍然适用.

多重比较 若零假设被拒绝，可对 2×2 列联表重复地使用中位数检验，对总体间进行逐对多重比较. 每次比较可以找到两个样本的中位数以及 2×2 列联表中大于或小于等于那个中位数的观测数. 在 2×2 列联表中用（1），（2）或（3）式计算检验统计量 T，若 $T>\chi^2_{1-\alpha}(1)$（自由度为 1 的 χ^2 分布随机变量的 $1-\alpha$ 分位数，可从表 A2

中获得），则认为这两个总体的中位数不相同.

计算机辅助 可在 *Minitab* 和 *StatXact* 中找到中位数检验. ◀

例 4.3.1

可用 4 种不同的方法来培植玉米，在被分割成若干块的土地上随机地采用这 4 种方
220 法并计算每块的亩产量.

方法

1	2	3	4
83	91	101	78
91	90	100	82
94	81	91	81
89	83	93	77
89	84	96	79
96	83	95	81
91	88	94	80
92	91		81
90	89		
	84		

为了决定产量差异是否由所用种植方法的不同而引起的，我们采用中位数检验，因为总体中位数的差异可以解释为所用种植方法的差异值. 假设表述如下：

H_0:所有种植方法有相同的亩产量中位数

H_1:至少有两种种植方法的亩产量中位数有差异

很容易算出这里有 34 个观测值，所以排序后的第 17 个和 18 个观测值的平均值，即 89 为总中位数. 将每种方法中大于 89 和小于或等于 89 的观测数记录如下：

	方法 1	2	3	4	总和
>89	6	3	7	0	16
≤89	3	7	0	8	18
总和	9	10	7	8	34

查表 A2，得临界域为 $T>\chi^2_{1-0.05}(c-1)=\chi^2_{0.95}(3)=7.815$（自由度为 $c-1=3$ 的 χ^2 随机变量的 0.95 分位数，可从表 A2 中获得）的区域. 用（1）式计算 T 为：

$$T=\frac{(34)^2}{(16)(18)}\left\{\frac{\left[6-\frac{(9)(16)}{34}\right]^2}{9}+\cdots+\frac{\left[0-\frac{(8)(16)}{34}\right]^2}{8}\right\}$$

221 $$=4.01(0.34+0.29+1.97+1.78)=17.6 \tag{4}$$

用（3）式可简便地计算 T，得：

$$T=\tfrac{9}{9}+\tfrac{16}{10}+\tfrac{49}{7}+\tfrac{64}{8}=17.6 \tag{5}$$

这与前面所得的值一样.

因为 17.6 的 T 值大于临界值 7.815，所以应拒绝 H_0，查表 A2 可知，p-值稍小于 0.001. 因为拒绝了 4 种玉米种植方法有相同中位数的零假设，所以采用多项逐对比较是合理的，将方法 1 与方法 2 比较，我们可以看到，19 个观测的样本中位数为 89，对方法 1，9 个观测值中有 6 个超过 89；对方法 2，10 个观测值中有 3 个超过 89，2×2 列联表检验统计量为 2.55，它小于 $\chi^2_{0.95}(1)=3.841$（自由度为 1 的 χ^2 随机变量的 0.95 分位数），因此，方法 1 和方法 2 的中位数不能认为不同. 然而，其他逐对比较在水平 $\alpha=0.05$ 下是显著的.

方法	中位数	T
1 和 2	89	2.55
1 和 3	92.5	6.35
1 和 4	83	13.43
2 和 3	91	13.25
2 和 4	82.5	14.40
3 和 4	82	15.00

■

通过在原始数据的子集上重复地使用相同的检验，除了第 1 对比较外，总体的多重比较方法往往会歪曲其他所有检验的显著水平. 这种重复的检验过程一般作为一种个人偏好或作为分离不同总体的一个客观“尺码”，但是整个检验中我们不能给出同第 1 对比较一样的合理解释. 关于重复检验过程的进一步探讨可参考 Gabriel (1966) 或 Knoke (1976).

例 4.3.1 中的试验是按照一种“完全随机化设计”来安排的，这种设计假设不同的方法是以某种随机方式（或等价于随机方式）分配到不同位置上的. 通常分析数据的参数方法被称为“单因素方差分析”，关于单因素方差分析的中位数检验的 A. R. E 依赖于总体分布函数的形式. 若总体为正态分布，则 A. R. E 仅为 $2/\pi=64\%$，然而，若总体是双指数分布（双指数分布拥有重尾），则 A. R. E 为 200%.

可将中位数检验推广到“分位数检验”，即零假设为“几个总体有相同的分位数”. 对任何选定的分位数，仅需改动检验的数据部分使得观测值被分为大于或小于等于整个数据排列的总分位数. 除了逼近式（3）不适用外，其余都同中位数检验一样，正如下面所给的理论，T 的精确分布（下面给出它的理论）也与中位数检验时有相同的形式. Wolfe (1977a) 给出了两阶段样本的中位数检验法. 222

□**理论** 正如上节第 3 种检验一样，因为一系列法则的目标是用于决定观测的计算是在列联表的上侧分位或下侧分位的格子里，所以行总和 a，b 是固定的. 例如，如果检验是一个“上侧四分位数”检验，则 a 大约为 $N/4$，b 大约为 $3N/4$，“大约”而不是“等于”是由于“允许样本中有结点或样本值相等”引起的，因此，T 的精确分布是依赖于行列总和的条件分布. 得到下列表（固定列总和）

$$\begin{array}{|c|c|c|c|}\hline O_{11} & O_{12} & \cdots & O_{1c} \\ \hline O_{21} & O_{22} & \cdots & O_{2c} \\ \hline \end{array} \begin{array}{c} a \\ b \end{array} \tag{6}$$
$$n_1 \quad n_2 \quad \cdots \quad n_c$$

的概率是二项分布的乘积

$$P\left(\begin{array}{|c|}\hline O_{1i} \\ \hline O_{2i} \\ \hline\end{array}\right) = \binom{n_i}{O_{1i}} p^{O_{1i}}(1-p)^{O_{2i}}; \qquad i = 1, 2, \ldots, c \tag{7}$$

这里 p 是一个观测大于总中位数的概率，现在 H_0 仅是表述"所有总体都拥有相同的中位数"，这并不意味着"所有总体都有相同的 p（超过样本总中位数的概率）". 另一方面，尽管这两种阐述（H_0 和前面的情形）在含义上很相似，但是从 H_0 不能推断出后一论断. 当 H_0 为真时，为了找到 T 的精确分布，我们必须要求对所有总体，超过总中位数的概率 p 相同，这就是为什么我们要给模型加第 4 个假设的原因.

因为样本彼此相互独立，用（7）式相乘，我们可以得出表（6）的联合概率分布

223

$$P\left(\begin{array}{|c|c|c|c|}\hline O_{11} & O_{12} & \cdots & O_{1c} \\ \hline O_{21} & O_{22} & \cdots & O_{2c} \\ \hline\end{array}\right) = \binom{n_1}{O_{11}}\binom{n_2}{O_{12}}\cdots\binom{n_c}{O_{1c}} p^{a}(1-p)^{b} \tag{8}$$

这里

$$a = O_{11} + O_{12} + \cdots + O_{1c}$$

和

$$b = O_{21} + O_{22} + \cdots + O_{2c}$$

给定了行总和 a，b，正如前一节的后面部分一样，用（8）式所得到的概率除以得到的行总和 a，b 的概率，则得（8）式中事件的概率，其结果是：

$$P\left(\begin{array}{|c|c|c|c|c}\hline O_{11} & O_{12} & \cdots & O_{1c} & a \\ \hline O_{21} & O_{22} & \cdots & O_{2c} & b \\ \hline n_1 & n_2 & \cdots & n_c & N \end{array}\right) = \frac{\binom{n_1}{O_{11}}\binom{n_2}{O_{12}}\cdots\binom{n_c}{O_{1c}}}{\binom{N}{a}} \tag{9}$$

与 4.2 节中（14）式相一致，我们不加推导地将（9）式按照多项系数记为如下形式：

$$\text{概率} = \frac{\begin{bmatrix} a \\ O_{1i} \end{bmatrix}\begin{bmatrix} b \\ O_{2i} \end{bmatrix}}{\begin{bmatrix} N \\ n_i \end{bmatrix}} \tag{10}$$

因此，T 的精确分布可以用（9）或（10）式求得，但我们很少用到，同上节一样（因为行数为 2），我们常用自由度为 $c-1$ 的 χ^2 分布来代替. □

▶中位数检验的推广

可将前面的中位数检验进行推广，使之能分析更复杂的试验，因为在中位数检验的推广中使用的记号很烦琐，我们通过一个应用例子来介绍这种检验. ◀

例 4.3.2

在6块不同的地里使用4种不同的肥料，并且整个试验重复使用3种不同的种子. 在试验的（4）（6）（3）=72个不同的条件下，得到试验的亩产结果如下： 224

		种子1				种子2 肥料				种子3			
		1	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
试验田	1	80.5	90.1	87.0	88.0	79.1	87.0	82.6	81.5	85.4	92.3	92.0	89.3
	2	87.0	83.4	89.1	90.3	77.6	82.0	81.4	87.9	89.2	90.1	90.2	93.6
	3	86.1	82.4	91.0	86.1	84.1	80.6	89.0	80.4	90.0	88.1	87.2	90.8
	4	82.1	84.9	84.4	83.1	83.3	79.5	86.3	83.1	83.4	85.3	94.3	87.6
	5	79.3	87.1	92.2	90.8	76.6	86.2	84.0	87.4	87.1	86.3	88.4	93.7
	6	84.2	89.3	85.3	84.7	81.0	84.1	88.1	85.0	82.3	92.9	95.1	82.9

为了检验零假设：

H_0:肥料不同不会引起产量中位数的差异

记第 i_2 块地上用肥料 i_1 和种子 i_3 观测到的产量为 $x_{i_1i_2i_3}$. 例如，x_{213}表示第1块地上使用肥料2和种子3后观测到的产量，从上表中可查到为92.3. 然后将 x_{213}与 $x_{113},x_{213},x_{313},x_{413}$的中位数相比较，后面的4个数代表了在其他因素等同的条件下，施不同的肥料得到的产量（H_0 说明施肥差异对产量没有影响）. 因此，x_{213}是同85.4,92.3,92.0和89.3的中位数

$$(\tfrac{1}{2})(89.3 + 92.0) = 90.65$$

进行比较，如果 x_{213}大于90.65，在表中由1代替，否则由0代替.

类似地，每一 $x_{i_1i_2i_3}$与 $x_{1i_2i_3},x_{2i_2i_3},\cdots,x_{ci_2i_3}$的中位数相比较，这些数代表了除了施肥的差异外，其他因素等同的条件下得到的产量. 在本例中，每个产量将与同一行（土地）同一块（种子）的中位数相比较，然后根据它们是否超过各自的中位数分别记为1或0. 结果如下：

		种子1				种子2 肥料				种子3			
		1	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
试验田	1	0	1	0	1	0	1	1	0	0	1	1	0
	2	0	0	1	1	0	1	0	1	0	0	1	1
	3	0	0	1	0	1	0	1	0	1	0	0	1
	4	0	1	1	0	1	0	1	0	0	0	1	1
	5	0	0	1	1	0	1	0	1	0	0	1	1
	6	0	1	1	0	0	0	1	1	0	1	1	0

记 O_j 为使用肥料 j 且得到的产量超过各自中位数的土地块数，即为上表在施肥料 j
225 条件下得到“1”的总数，下表给出所有的 $O_j, j=1,2,\cdots,c$.

	肥料 1	肥料 2	肥料 3	肥料 4	总和
O_j =“1”的数目	3	8	14	10	$a = 35$
“0”的数目	15	10	4	8	$b = 37$
	$n_1 = 18$	$n_2 = 18$	$n_3 = 18$	$n_4 = 18$	$N = 72$

则可将通常的中位数检验用于此表. 由（3）式，我们得到：

$$T = \frac{(144 + 4 + 100 + 4)}{18} = 14.0 \tag{11}$$

将 T 与 $\chi^2_{0.95}(3)$（自由度为 $c-1=3$ 的 χ^2 随机变量的 0.95 分位数）相比较，查表 A2 得 $\chi^2_{0.95}(3)=7.815$，由于 $T>7.815$，所以应拒绝 H_0，相应的 p-值大约为 0.004.

■

习题

1. 检验零假设：以下样本来自于中位数相同的总体

样本 1: 35, 42, 42, 30, 15, 31, 29, 29, 17, 21
样本 2: 34, 38, 26, 17, 42, 28, 35, 33, 16, 40
样本 3: 17, 29, 30, 36, 41, 30, 31, 23, 38, 30
样本 4: 39, 34, 22, 27, 42, 33, 24, 36, 29, 25

2. 将一些油井拍卖给出价最高的投标人，每个油井都收到至少一个密封投标. 检验零假设：已经投入生产的油井与还未投入生产的新井有相同的中位投标数. 从两种类型的油井中各抽取一个随机样本，结果如下：

	每个油井的投标数
投入生产的	6, 3, 1, 14, 8 9, 12, 1, 3, 2, 1, 7
未投入生产的	6, 2, 1, 1, 3, 1, 2, 4, 8, 1, 2

3. 例 4.3.2 中的试验结果是否说明种子差异显著?
4. 例 4.3.2 中的试验结果是否说明土地差异显著?

226 5. 30 种股票的随机样本分别选自美国三大股票交易市场，并记录下它们上一年的运作情况. 整理所有 90 种股票的中位数运作记录，并列表如下：

交易市场	超过中位数的股票数
纽约	18
美国	17
纳斯达克	10

上一年三个交易市场的股票运作是否有显著差异?

6. 随机地将 100 名入伍新兵分配给海军训练新兵营的 4 位军士来训练，训练将结束时，84 名新兵留了下来，并且记录了他们障碍课程的成绩.

军士	训练新兵数	成绩超过中位数的士兵人数
Adams	20	11
Baker	22	8
Callahan	20	8
Davis	22	15

4 位军士的训练结果是否有显著差异?

思考题

1. 若 $a=b$，证明（1）式可变成（3）式.
2. 若 $r=2$，证明（1）式与（4.2.11）式相同.
3. 本节中设计的常用参数检验（单因素情形）通常假设每个观测服从正态分布，而不是仅仅依赖低于或高于中位数所得到的 0 或 1. 若每个样本中的观测当它们低于总中位数时称为 0 集，大于或等于总中位数时称为 1 集，通过计算 0 集和 1 集上的个数，前面参数检验的统计量可简化为：

$$F=\frac{\left(\sum_{j=1}^{c}\frac{O_{1j}^{2}}{n_j}-\frac{a^2}{N}\right)(N-c)}{\left(a-\sum_{j=1}^{c}\frac{O_{1j}^{2}}{n_j}\right)(c-1)}$$

证明 F 可以写成如下 T 的函数

$$F=\frac{T(N-c)}{(N-T)(c-1)}$$

因此，当 T 值较大时拒绝 H_0 与当 F 值较大时拒绝 H_0 等价.

4.4　相依性度量

列联表可以方便地检验数据是否具有某种内在的相依性. 通过列联表，可以揭示一种特殊类型的相依：行—列相依. 若不同的行代表从不同总体中抽取的样 227
本，不同的列代表同一样本中数据不同的分类，则一个行—列相依与所抽取样本的总体上不同类中的概率函数相依同义. 类似地，如果某个随机样本中的观测根据两个不同准则被归入行与列中，显然一个行—列相依可解释为分类两个准则间的一种关系.

正如本章中到目前为止我们所做的，代替检验假设，我们仅希望表达一个给定列联表所示的相依度. 当然，我们更愿意以一种简单的形式，以及把由列联表表现出的精确相依度易于传授给他人的形式来表示相依度.

与第 1 种方法一样，我们可用上节的检验统计量

$$T=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \tag{1}$$

作为相依性的度量，一般来说，“如果它对相依性检验足够好，那么对相依性的度量也就足够好.” T 的使用似乎满足我们所考虑的方便和简捷，然而，为了将相依度传授给其他人，应标明 T 的自由度，这是因为没有自由度就不可能告诉别人 T 值所揭示的相依度. 即使知道自由度，为了解释 T，非专业人士也必须借助于一张 χ^2 表.

例 4.4.1

例 4.2.1 中的列联表如下：

	分数				
	0-275	276-350	351-425	426-500	总和
私立学校	6	14	17	9	46
公立学校	30	32	17	3	82
总和	36	46	34	12	128

228 计算本列联表的 T 为 17.3，现在 17.3 近似为 $\chi^2_{0.999}(3)$，所以

$$p\text{-值}=1-p=0.001$$

如此小的 p-值说明数据强烈不“同意”行分类（学校类型）与列分类（测验分数）独立的零假设，但是它并不能度量相关水平. ■

▶Cramér 关联系数

一种易于解释的相依性度量方法就是包括修改（1）式中的 T，使得结果不像 T 一样太多地依赖于自由度. 一种修改是用 T 除以达到可能的最大 T 值，目前我们知道大值 T 来自格子计数显著不平衡的列联表，通过检测极端不平衡的列联表（假设样本容量为 N，给定 r 行 c 列），我们由基本试验和误差发现，当每行每列中除了一个格子里的数外其余全为 0（如果 $r\neq c$，一些行或列全为 0）时 T 最大. 即有 T 在下列列联表中达到最大值

	列					
	1	2	3	4	5	总和
行 1	3	0	0	0	0	3
2	0	3	0	0	0	3
3	0	0	3	0	0	3
总和	3	3	3	0	0	9

在定义 0/0 = 0 或等价地忽略行或列全为 0 的格子后，此表中 $T = 18$.

一般情况下，T 的最大值为 $N(q-1)$，这里 q 为 r,c 中较小的一个，N 为观测总数. T 除以最大数后为：

$$R_1 = \frac{T}{N(q-1)} \tag{2}$$ 229

其中 q 为 r,c 中较小者. 若列联表中指明一个强行—列相依性时，则 R_1 接近于 1.0，若列联表中每行中的列数相互有相同的比例，且这个比例与列总和间的相互比例相同，则 R_1 接近于 0. 这种度量是由 Cramér(1946, p. 443)提出的，R_1 的方根即为“Cramér 系数”，很多现代的计算包，比如 *SAS* 和 *StatXact* 都可以计算.

$$\text{Cramér系数} = \sqrt{\frac{T}{N(q-1)}} \tag{3}$$

这也是目前度量 $r \times c$ 列联表相依性最广泛使用的方法. ◀

例 4.4.2

在上例 2 × 4 列联表中，$T = 17.3$，因为 $N = 128$ 和 $q = 2$，我们由（2）式，得到

$$R_1 = \frac{T}{N(q-1)} = \frac{17.3}{128} = 0.135$$

和 Cramér 系数为 $\sqrt{0.135} = 0.368$. ■

像所有好的相依性度量一样，Cramér 系数是“尺度不变”的，即如果试验数值的尺度变大，比如例 4.4.1 中将学生人数扩大 10 倍，只要所有观测值变化一致，则相依性的度量不会改变. 如果 10 倍多的学生参与测试，则每格中的观测结果也扩大 10 倍，相应的列联表结果如下：

	分数				
	0–275	276–350	351–425	426–500	总和
私立学校	60	140	170	90	460
公立学校	300	320	170	30	820
总和	360	460	340	120	1280

检验统计量 T 也扩大了 10 倍，为 173，但 Cramér 系数仍为 0.368. 这是因为对较小尺度试验的相依度同较大尺度试验的相依度相同. 230

▶Pearson 关联系数

除 Cramér 系数外，另外两种关联系数有时也用到. 第一种是均方关联 *Pearson* 系数（Pearson's coefficient of mean square contingency），它由 Yule 和 Kendall(1950, p. 53)给出，定义为：

$$R_2 = \sqrt{\frac{T}{N+T}} \tag{4}$$

我们已说明 T 的最大值为 $N(q-1)$，所以 R_2 的最大值为：

$$R_2(\max) = \sqrt{\frac{N(q-1)}{N+N(q-1)}} = \sqrt{\frac{q-1}{q}} \tag{5}$$

在很多情形下，它接近于 1，T 的最小值可能为 0，所以

$$0 \leqslant R_2 \leqslant \sqrt{\frac{q-1}{q}} < 1.0 \tag{6}$$

McNemar(1962, p. 198)和 Siegel (1956, p. 196)也称 R_2 为关联系数 (contingency coefficient)． ◀

例 4.4.3

在前面两例的列联表中，我们有 $T = 17.3$ 和 $N = 128$，所以

$$R_2 = \sqrt{\frac{T}{N+T}} = \sqrt{\frac{17.3}{128+17.3}} = 0.345$$

■

▶ Pearson 均方关联系数

我们给出第 3 种相依性度量 R_3，它也具有 Pearson 关联系数的特点(见 Cramér, 1946, p. 282)，也被 Yule 和 Kendall(1950, p. 53)称为 Pearson 均方关联系数 (mean-square contingency coefficient)．R_3 的定义如下：

$$R_3 = \frac{T}{N} \tag{7}$$

由上述讨论，我们得到：

231

$$0 \leqslant R_3 \leqslant q-1$$

为了正确解释由 R_3 得到的相依度，我们需要知道 r，c 的知识． ◀

例 4.4.4

用前例中的列联表，我们得到

$$R_3 = \frac{17.3}{128} = 0.135$$

■

最后，我们简单提一下 *TschuProw* 系数 (Tschuprow's coefficient)，它由 Yule 和 Kendall(1950)提出，形式如

$$R_4 = \sqrt{\frac{T}{N\sqrt{(r-1)(c-1)}}} \tag{8}$$

相依性度量方法的选择很大程度上取决于个人决定，最初的动机来自于传统习惯，而不是统计上的考虑．进一步的讨论见 Stuart(1953)．

对于 2×2 列联表，

	列 1	列 2	
行 1	a	b	r_1
行 2	c	d	r_2
	c_1	c_2	N

前面的度量可以简化一些. 从习题4.2.2，我们知道 T 可写成：

$$T=\frac{N(ad-bc)^2}{r_1r_2c_1c_2} \tag{9}$$

因此 R_3，R_1（因为 $q=2$）简化为

$$R_1=R_3=\frac{T}{N}=\frac{(ad-bc)^2}{r_1r_2c_1c_2} \tag{10}$$ 232

并且 Cramér 系数变为

$$\sqrt{R_1}=\sqrt{\frac{T}{N(q-1)}}=\sqrt{\frac{(ad-bc)^2}{r_1r_2c_1c_2}} \tag{11}$$

R_2 可写为

$$R_2=\sqrt{\frac{T}{N+T}}=\sqrt{\frac{(ab-bc)^2}{r_1r_2c_1c_2+(ab-bc)^2}} \tag{12}$$

不同于 $r\times c$ 列联表，在一个 2×2 列联表中，有时区分正关联和负关联是有意义的，比如当分类的两个标准有对应的类型时.

例 4.4.5

根据母亲的头发和父亲的头发是否为黑色或金色将 40 个孩子分类，根据 $ad-bc$ 是正或负，则结果可能表现出正关联（正相关）

		父亲 黑色	父亲 金色	
母亲	黑色	28	0	28
母亲	金色	5	7	12
		33	7	40

或负关联（负相关）

		父亲 黑色	父亲 金色	
母亲	黑色	21	7	28
母亲	金色	12	0	12
		33	7	40

233

无关联（零相关）的情形可表示如下：

		父亲		
		黑色	金色	
母亲	黑色	23	5	28
	金色	10	2	12
		33	7	40

■

▶**Phi 系数**

如果要找到关联的类型（正，负或零相关），必须注意建立列联表使得 a 和 d 能代表相似分类（黑色—黑色，金色—金色）数，而 b 和 c 能代表相异分类（黑色—金色，金色—黑色）数. 一种保留方向关联类型的度量是 Phi 系数（phi coefficient），由下式给出：

$$R_5=\frac{ad-bc}{\sqrt{r_1r_2c_1c_2}} \tag{13}$$

它的范围为 -1 到 +1，其中 -1 表示所有对象都被归入到“相异”类中($a=d=0$)，而 1 表示所有对象都被归入到“相似”类中（$b=c=0$）. 其实 Phi 系数只不过是保留了符号（即 $ad-bc$ 的符号）的 Cramér 系数（见（11）式）. Phi 系数方法被广泛采用的一个原因是，它是 Pearson 乘积矩相关系数（见下章）的特殊情形，它是用数来代表类进行计算的. 注意，在 2×2 列联表分析中使用的检验统计量 T_1 与 Phi 系数间有着紧密关系，(4.1.1) 式表明 Phi 系数等于 $T_1/\sqrt{N}$. ◀

例 4.4.6

对例 4.4.5 中的第 1 个表，我们有

$a=28 \quad r_1=28 \quad b=0 \quad r_2=12$

234 $c=5 \quad c_1=33 \quad d=7 \quad c_2=7$

所以，R_5 可如下计算：

$$R_5=\frac{ad-bc}{\sqrt{r_1r_2c_1c_2}}=\frac{(28)(7)-0}{\sqrt{(28)(12)(33)(7)}}=0.703 \tag{14}$$

对例 4.4.5 中的第 2 个表，计算 R_5 为

$$R_5=\frac{(21)(0)-(7)(12)}{\sqrt{(28)(12)(33)(7)}}=-0.302 \tag{15}$$

它反映了头发类型的负关联. ■

其他 2×2 列联表的关联性度量方法还包括由 Yule 和 Kendall(1950, p. 30)提出的

$$R_6=\frac{ad-bc}{ad+bc} \tag{16}$$

与 Ives 和 Gibbons（1967）提出的

$$R_7=\frac{(a+d)-(b+c)}{a+b+c+d} \tag{17}$$

可定义的关联度量方法有许许多多，选择何种系数取决于个人偏好.

有时我们面临这样的问题："我们如何用 R_1（或 R_2, R_3 等）作为检验统计量来检验具有独立性的零假设."答案是：用求 4.2 节中 T 的精确分布同样的步骤，你可以找到任意一个度量的精确小样本分布．因此理论上总可以设计一个检验，但事实上对同样假设检验，它比 4.1 节和 4.2 节来得更容易且更有效.

特别地，系数

$$R_1 = \frac{T}{N(q-1)}$$

$$R_2 = \sqrt{\frac{T}{N+1}}$$

$$R_3 = \frac{T}{N}$$

和

$$R_4 = \sqrt{\frac{T}{N\sqrt{(r-1)(c-1)}}}$$ 235

同 Cramér 系数一样，当 T"太大"时，它们也会"太大"，原因是它们随着 T 的增减而增减．在 4.2 节的检验中，我们用 T 作为检验统计量，当 T 显著时，可以推出 R 显著.

基于 Phi 系数 R_5 的单边检验对 2×2 列联表情形是合适的，原因是它们有如下关系：

$$R_5 = \frac{T_1}{\sqrt{N}} \tag{18}$$

这里的 T_1 由（4.1.1）式所给出，且 T_1 是渐近正态的，因此 $\sqrt{N}\cdot R_5$ 也是渐近正态的. 所以若 $\sqrt{N}\cdot R_5$ 太大（对于水平 α，即 $\sqrt{N}\cdot R_5$ 超过了 $z_{1-\alpha}$），就应拒绝零假设

H_0:不存在正相关

并且若 $\sqrt{N}\cdot R_5$ 太小（对于水平 α，即 $\sqrt{N}\cdot R_5$ 小于 z_α），就应拒绝零假设

H_0:不存在负相关

这与 4.1 节中所描述的基于 T_1 的检验完全相同.

例 4.4.7

为了检验安全带能否防止死亡事故，我们提取了发生在高速公路上的 100 起车祸的记录做研究，这 100 起事故涉及到 242 个人，每个人根据事故发生时是否系安全带和是否带来致命伤害来分类，结果如下：

		致命伤害?		
		是	否	总和
系安全带	是	7	89	96
	否	24	122	146
	总和	31	211	242

我们希望证明的是："安全带可以防止死亡事故的发生." 然而，一个相关检验不能
236 自动地给出因果关系，而两个变量间的因果关系往往导致相关，一个显著相关又有可能是这两个变量都会被第三个变量影响，如本例，两个变量都可以受到司机的粗心大意这个变量的影响. 因此零假设是

H_0:在一次交通事故中，系安全带与死亡事故不存在负关联备择假设为

H_1:在一次交通事故中，系安全带与死亡事故存在负关联

本例与我们曾描述的两准则分类没有相同对应类的情形有少许不同（"是"和"否"意味着在行中是一类，在列中又是另一类），因此我们需要考虑所面临的是哪一种情形. 本例中，若 H_1 为真，则我们期望 b 和 c 大于 a 和 d，因此不等式

$$ad - bc < 0$$

趋于支持 H_1，从而导致 R_5 为负. 所以若 $\sqrt{N}\cdot R_5$ 小于 -1.645（标准正态分布的0.05分位数，由表A1获得），则拒绝 H_0，本例中的检验统计量：

$$\sqrt{N}\cdot R_5 = \frac{\sqrt{N}(ad-bc)}{\sqrt{r_1 r_2 c_1 c_2}} = \frac{\sqrt{242}[(7)(122)-(89)(24)]}{\sqrt{(96)(146)(31)(211)}} = -2.0829$$

小于 -1.645，所以应拒绝 H_0. 我们可以得出使用安全带与减少死亡事故是有关联的（这种关系是否为因果关系仍是一个公开的问题）. 同时可从表A1中查到 p-值约为0.019. ■

其他几种基于两准则分类得到变量间的相依性度量方法可参考 Goodman 和 Kruskal(1954,1959,1963)的经典文章. Davis（1967）介绍了偏系数方法.

习题

237 1. 调查100对已婚夫妇，分别询问丈夫和妻子希望谁是下一届美国总统的首选，结果如下：

		妻子的选择		
		A	B	其他
丈夫的选择	A	12	22	6
	B	25	21	4
	其他	3	7	0

计算：

(a) T　　(b) Cramér 系数　　(c) R_1

(d) R_2　　(e) R_3　　(f) R_4

2. 一名护士让50位遭受关节炎折磨的工人服下两种不同的药物，其中25人服下阿司匹林，另外25人在不知情的情况下服下安慰剂. 一个小时后，询问工人药物是否让他们感觉好些，服阿司匹林组有17人和服安慰剂组有12人给出肯定回答.

(a) 使用 R_5 来检验服用阿司匹林与"感觉好些"间是否存在正相关.

(b) 计算 R_6. (c) 计算 R_7.

3. 对一个交通状况良好的城市街道进行短期的交通状况调查，共调查了 64 辆小轿车，其中 16 辆超速，而 48 辆没有，并且 24 辆带有乘客而其余车上只有司机一人，12 个超速者独自驾车. 假设观测的交通状况是整个交通状况的一个随机样本，则

(a) 使用 R_5 来检验超速与独自驾车二者之间是否存在正相关.

(b) 计算 R_6. (c) 计算 R_7.

4. 在美国西南部的湖中发现了某种昆虫，对这种昆虫的一项研究就是为了找出这种昆虫的染色体结构是否会因为所在州的不同而有显著差异. 不同染色体类型的昆虫数记录如下：

类型	得克萨斯	新墨西哥	亚利桑那	加利福尼亚
A	54	72	83	96
B	20	6	18	6
C	17	3	12	0
D	0	12	14	1
E	0	10	0	0

计算：

(a) T (b) Cramér 系数 (c) R_1

(d) R_2 (e) R_3 (f) R_4 238

思考题

1. 在下面的列联表中，试证明：$T=N(q-1)$（这里 $r<c$）.

	1	2	···	r	r+1	···	c
1	O_{11}	0	···	0	0	···	0
2	0	O_{22}	···	0	0	···	0
···	···	···	···	···	···	···	···
r	0	0	···	O_{rr}	0	···	0

2. 对于习题 1，考虑另外一个 $r\times c$ 列联表，使你怀疑 T 有较大的值，并计算你构造的列联表的 T 值，它是否大于 $N(q-1)$？

3. 证明如下等式：

(a) $R_2^2=\dfrac{R_3}{1+R_3}$.

(b) $R_3=\dfrac{R_2^2}{1-R_2^2}$.

(c) 当 $r=2$ 和 $c=2$ 时，$R_3=R_5^2=R_1$.

4. 证明 Phi 系数是数对 (X_i,Y_i) 上的 Pearson 乘积矩相关系数：

$$r=\frac{\sum_{i=1}^{N}(X_i-\overline{X})(Y_i-\overline{Y})}{\left[\sum_{i=1}^{N}(X_i-\overline{X})^2\sum_{j=1}^{N}(Y_j-\overline{Y})^2\right]^{1/2}}$$

其中，(X_i,Y_i) 根据观测所在格分别取值 (0,0)，(0,1) (1,0) 或 (1,1)，根据上面的公式计算相关系数，然后说明 0 和 1 两个数可由任意两个数 p 和 q 代替，其结果也成立.

4.5 χ^2 拟合优度检验

人们通常关心被观测随机变量的未知概率分布的假设检验问题，例如："中位数
239 为4.0" 和 "两个总体在类1中的概率一样". 我们遇到更加详尽的假设包括："未知分布函数是均值为3.0，方差为1.0的正态分布函数" 或 "随机变量服从参数为 $n=10$ 和 $p=2$ 的二项分布". 后面这两个假设更加详尽是因为这些假设不仅仅是关于概率某些方面的陈述，比如中位数，而是关于整个概率分布的陈述，因此也就包括了所有概率和分位数的陈述. 后两种类型的假设可以用 "拟合优度检验" 来检验，即设计一个检验来比较从假设的分布中抽取的样本，看所假设的分布函数与样本数据是否 "拟合".

最悠久和众所周知的拟合优度检验是 χ^2 拟合优度检验，它由 Pearson(1900) 首次提出.

▶ χ^2 拟合优度检验

数据 数据由随机变量 X 的 N 个观测组成，这 N 个观测划分为 c 类，并且在每类中的观测数可归入到如下 $1\times c$ 列联表中.

	类 1	2	···	c	总和
观测频数	O_1	O_2	···	O_c	N

记类 O_j 为类 j 中的观测数，$j=1,2,\cdots,c$.

假定条件

1. 样本是随机的.
2. 度量尺度至少为名义的.

检验统计量

在零假设 H_0 为真的条件下，令 X 的一个随机观测落入类 j 的概率为 p_j^*. 定义 E_j 为

240
$$E_j=p_j^*N,\qquad j=1,2,...,c \tag{1}$$

这里 E_j 表示 H_0 为真时，观测落入类 j 的期望观测数. 给出如下检验统计量 T:

$$T=\sum_{j=1}^{c}\frac{(O_j-E_j)^2}{E_j} \tag{2}$$

一个计算方便而等价的表达式为

$$T=\sum_{j=1}^{c}\frac{O_j^2}{E_j}-N \tag{3}$$

零分布 T 的精确分布难以求得，所以我们用自由度为 $c-1$ 的χ^2 分布来近似.

假设

$$H_0: P\ (X\text{在类}\ j\ \text{中})\ =p_j^*, j=1,2,\cdots,c$$

$$H_1: P\ (X\text{在类}\ j\ \text{中})\ \neq p_j^*,\ \text{对某个}\ j$$

若 $T>\chi^2_{1-\alpha}(c-1)$（自由度为 $c-1$ 的χ^2 分布的 $1-\alpha$ 分位数，由表 A2 获得），则拒绝 H_0，p-值近似等于 $P(\chi^2(c-1)>T)$，这个概率可由表 A2 获得.

计算机辅助 可在 *Minitab*, *S-plus* 和 *StatXact* 中运行χ^2 拟合优度检验. ————◀

评注

如果一些 E_j 太小，用χ^2 分布渐近（在下面描述）可能不合适，但究竟小到什么程度可能还不清楚. 但是，Cochran(1952)建议所有 E_j 都不能小于 1 且有不超过 20% 的 E_j 小于 5. 最近的研究结果表明这个限制还可以再放宽一些，Yarnold(1970)提到："如果类的数目 s 为 3 个或 3 个以上，r 为类中期望观测数小于 5 的数目，则最小的期望观测数可以小到 $5r/s$." Slakter(1973)认为类的数目可以超过观测数，这意味着平均期望值可以小于 1. 更近一些的研究成果是由 Koehler 和 Larntz(1980)发现的，他们指出，只要 $N\geqslant 10, c\geqslant 3, N^2/c\geqslant 10$ 并且所有 $E_j\geqslant 0.25$，用χ^2 分布来近似都是合适的. 如果有很多个 E_j 偏小，使用者可以基于 Koehler 和 Carrotz 的研究考虑合并一些格子. 241

例 4.5.1

用某种计算机程序来产生随机个位数，如果程序运行正常，则计算机将输出数字 (2,3,7,4 等)，这些数字可看作独立同分布随机变量的观测，这里每个数字 0,1,2,…,8,9 是等可能得到的（概率为 0.1），我们要检验

H_0:数字是随机出现的

对备择假设：

H_1:某些数字较其他数字更可能出现

则一种方法就是计算每个数字出现的次数，以下是产生的 300 个数字.

1578748416	4705188926	6936349612
4653843213	0282868892	3928057043
5101259393	9837006785	3011679938
7122863085	6528271107	2956427027
2671728075	9759178719	9373309535
8363265100	2546793732	2212122529
9453087720	3976759377	9593511031
5605373242	1819898287	3872181027
3494768396	9296177240	8620774591
4659773922	9246724287	8326143939

在零假设下，每个数字是等可能出现的，所以0,1,2,…,9中每个数字出现的期望次数应为30，但是数字2出现了41次而数字4仅出现了19次，这是不是由随机波动引起的呢？

完整的观测计数如下：

数字	0	1	2	3	4	5	6	7	8	9	总和
观测频数	22	28	41	35	19	25	25	40	30	35	300
期望频数	30	30	30	30	30	30	30	30	30	30	300

检验统计量为：

$$T=\sum_{i=1}^{10}\frac{O_i^2}{E_i}-N=317-300=17 \tag{4}$$

242 因为 $T>\chi^2_{0.95}(9)=16.92$（自由度为9的 χ^2 分布的0.95分位数，由表A2获得），因此在置信性水平0.05下拒绝零假设，即更倾向于备择假设，说明数字不是等可能由计算机程序产生的. 查表可知，p-值略小于0.05. ■

评注

如果 X 的概率分布除了 k 个参数外完全确定，则应首先估计这 k 个参数，然后再进行拟合优度检验. 惟一的变化是此时 T 的渐近 χ^2 分布的自由度为 $c-1-k$，也就是说，估计一个参数，自由度应减去1. 然而只有参数以某种合适的方式估计时，自由度才可以减去1. 例如，在一个分成4类的拟合优度检验中，若 $T>7.815$（见表A2），通常应拒绝 H_0（在 $\alpha=0.05$）. 然而，如果在检验前就用数据估计一个参数，则修正的假设分布可以更好地同数据拟合.（如果这个估计是一个“好”的估计，则修正后的假设分布会同数据拟合得很好. 一个不好的估计会导致分布与数据拟合得不好，那么拟合优度检验可能不再有效. Chase(1972)讨论了当独立于数据估计参数时的 χ^2 检验.）

拟合优度检验倾向于保护 H_0，所以检验比较保守，且检验的功效不高. 我们希望扩大临界域使得 $\alpha=0.05$，并找回检验失去的功效. 若我们减去1个自由度，如用2个自由度代替3，就可以扩大临界域，这时当 $T>\chi^2_{0.95}(2)=5.991$（而不是 $\chi^2_{0.95}(3)=7.815$）时，拒绝 H_0. 问题是：“能证明我们减去1个自由度的做法合理吗？”

Cramér(1946,p.424或见Birnbaum,1962,p.258)证明如果1个参数由最小 χ^2 方法估计所得，则可以减去1个自由度. 最小 χ^2 方法就是在 χ^2 检验统计量中给定观测值使统计量达到最小的参数值来估计参数. 实际操作中就是将参数的所有可能值，或几个未知参数的所有可能组合带入公式中计算 E_j 和 T，然后找到使 T 最小的参数值. 然而，这个过程是繁琐的，因此Cramér提出了一种更可行的修正最小 χ^2 方法(modified minimum chi-squared method)，但这仍然很繁琐. 所以Cramér和Birnbaum在他们给出的例子中，实际用的是修正最小 χ^2 方法的一个修正，如果使用这种估计

方法来估计一个参数，则渐进地允许减去 1 个自由度. 实际参数估计方法是通过计算分组数据的前 k 阶样本矩来估计 k 个参数，（假设每个观测位于所在组区间的中点处，所有包含观测的区间长都有限.）然后令前 k 阶总体矩与分组数据的前 k 阶样本矩相等，由 k 个联立方程解出 k 个参数. 下面的例子和后面的评注可以帮助我们弄明白上述方法. 243

例 4.5.2

Efron 和 Morris(1975)提供了 1970 年大联盟棒球队排在前 18 名的运动员 45 次出棒的记录. 运动员的名字和他们 45 次出棒击中的次数如下：

Clemente	18	Kessinger	13	Scott	10
F. Robinson	17	L. Alvarado	12	Petrocelli	10
F. Howard	16	Santo	11	E. Rodriguez	10
Johnstone	15	Swoboda	11	Campaneris	9
Berry	14	Unser	10	Munson	8
Spencer	14	Williams	10	Alvis	7

我们将检验零假设：这些数据服从 $n=45$ 的二项分布，但我们首先需要估计每次出棒击中的概率 $p=P$（击中）.

对于 p，一个好的估计是用这些数据中击中的相对频数来估计.

$$\hat{p}=\text{总击中次数}/\text{总出棒次数}=\frac{215}{810}=0.2654 \tag{5}$$

然后计算 $n=45$，$p=0.2654$ 的二项分布的概率，

$$P(X=i)=\binom{45}{i}(0.2654)^i(0.7346)^{45-i} \qquad i=0,1,...,45 \tag{6}$$

每格期望的观测数为

$$E_i=18\cdot P(X=i) \qquad i=0,1,...,45 \tag{7}$$

将期望观测数小于 0.5 的格子合并起来，这样做是统计量 T 的分布有较好的 χ^2 分布逼近. 合并后的结果如下：

	击中数												
	<7	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**>18**	**总和**
观测值	1	1	1	5	2	1	1	2	1	1	1	1	18
期望值	1.10	1.06	1.57	2.04	2.35	2.40	2.20	1.82	1.36	0.92	0.57	0.61	18

检验统计量为

$$T=\sum_{i=1}^{12}\frac{O_i^2}{E_i}-N=24.73-18=6.73 \tag{8}$$ 244

因为 $T<\chi^2_{0.95}(12-1-1)=\chi^2_{0.95}(10)=18.31$（自由度为 10 的 χ^2 分布的 0.95 分位数，由表 A2 获得），所以在 $\alpha=0.05$ 的水平下接受 H_0. 事实上，T 的观测值为 6.73，由表 A2 可得，p-值大于 0.25，所以二项分布同数据拟合得相当好. 注意，这里我们为什么要将自由度减去一个 1，是因为参数 p 是用数据估计的. ■

评注

例4.5.2中用18个与动员的总击中次数/总出棒次数来估计参数 p，这个估计相当好，但不一定就是使得 T 达到最小的估计. 与渐近理论一致，因为估计了一个参数 p，所以自由度应减去1. 这里需要注意两点.

第一，例4.5.2中 p-值大大超过0.25，所以零假设很容易接受，这里没有必要去找 T 的最小值来进一步提高 p-值. 结论还是一样，即二项分布同数据拟合得相当好，因此除非 p-值相当小且对结论有所怀疑外，没有必要再找最小 χ^2 统计量.

第二，用最小 χ^2 方法估计参数时，判断是否减去一个自由度的理论实际上是一种渐近论（样本趋于无穷，且每个单元格的期望观测数也趋于无穷）. 而对小样本情形，用最小 χ^2 方法不能保证能得到充分的近似，但实际生活中我们常碰到小样本，因此我们可能用传统的估计方法来估计未知参数更牢靠，比如矩估计或极大似然估计，或者我们知道对被检样本用最小 χ^2 方法时，χ^2 近似也比较好. 关于这个论题更详尽的论述见 Yule 和 Kendall(1950)，Chernoff 和 Lehmann(1954)及 Berkson(1980).

前面两个例子中所讨论的随机变量都是离散型随机变量，事实上，χ^2 拟合优度检验并不限于离散型随机变量情形，它也可以检验数据是否来自于某一指定的连续分布，并且同例4.5.2一样，其中的某些未知参数可通过数据来估计. 第一步通过分区间来离散化连续型随机变量，这些区间即为检验中的分类. 当零假设为真时，每一区间中的观测数 O_j 将与期望的观测数

$$E_j = N \cdot P(X\text{在区间}j\text{中}) \tag{9}$$

245 进行比较.

下面的例子帮助我们了解，如何对含有两个未知参数的连续分布进行拟合优度检验. 注意，如何分区间比较主观，因此这是用 χ^2 优度检验来检验连续型分布时的一个薄弱环节.

例4.5.3

从一个电话簿中随机抽取50个两位数，并用 χ^2 拟合优度检验来检验这50个观测是否来自于服从正态分布的总体. 将数据按照升序排列如下：

23	23	24	27	29	31	32	33	33	35
36	37	40	42	43	43	44	45	48	48
54	54	56	57	57	58	58	58	58	59
61	61	62	63	64	65	66	68	68	70
73	73	74	75	77	81	87	89	93	97

零假设是：

H_0:这些观测来自于正态分布的随机变量

正态分布有两个参数（见定义1.5.3），但零假设中并没有确定它们的具体值，因此在用拟合优度检验前需要把它们估计出来，整个过程分为如下几步.

第一步 将观测分别归入长度有限的几个区间中. 我们可任意选择区间，比如20—40，40—60，60—80和80—100，其中不包括每个区间的上限.

	区间段 20 — 40	40 — 60	60 — 80	80 — 100	总和
观测数	12	18	15	5	50

第二步 用分组数据的样本均值 $\overline{X}$ 和样本标准差 S 来估计 μ 和 σ. 将区间段 20—40 的 12 个观测都看作这段区间的中点值 30，40—60 的 18 个观测都等于 50，等等，然后用这些数按照定义 2.2.3 的公式算出样本均值 $\overline{X}$ 和样本标准差 S： 246

$$\overline{X}=\frac{1}{N}\sum_{i=1}^{N}X_i$$

$$=\frac{1}{50}[12(30)+18(50)+15(70)+5(90)]=55.2 \tag{10}$$

$$S=\sqrt{S^2}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}X_i^2-\overline{X}^2}$$

$$=\left\{\frac{1}{50}[12(30)^2+18(50)^2+15(70)^2+5(90)^2]-(55.2)^2\right\}^{1/2}=18.7 \tag{11}$$

因此，μ 和 σ 的估计分别为 55.2 和 18.7.

第三步 用第二步得到的 μ 和 σ 的估计，计算第一步中的所有区间段的 E_j 和尾部概率.

类的边界 b_j	$(b_j-\overline{X})/S=x_p$	$F(x_p)$	区间段	p_j^*
$b_1=20$	−1.88	0.03	<20	0.03
$b_2=40$	−0.813	0.21	20 − 40	0.18
$b_3=60$	+0.256	0.60	40 − 60	0.39
$b_4=80$	+1.33	0.91	60 − 80	0.31
$b_5=100$	+2.40	0.99	80 − 100	0.08
			≥100	0.01

当假设分布是均值为 55.2，标准差为 18.7 的正态分布时，为了找到观测在不同类中的假定概率，我们认为类的边界（表中的第一列）是假定分布的分位数. 这些分位数可通过公式 1.5.3 转化为服从标准正态分布的随机变量（第二列）的分位数，通过查表找出边界代表的是哪个分位数（第三列），第三列中的后一项减去前一项就可以得出观测在假设分布下落入每区间段的概率 p_j^*. 由（1）式，$E_j=50p_j^*$，列表如下：

	类 <20	20−40	40−60	60−80	80−100	≥100
期望数 E_j	1.5	9.0	19.5	15.5	4	0.5
观测数 O_j	0	12	18	15	5	0

247

由于一些 E_j 太小，将第一个和最后一个单元格合并成如下表：

	类			
	<40	40–60	60–80	≥80
期望数 E_j	10.5	19.5	15.5	4.5
观测数 O_j	12	18	15	5

第四步 计算 T. 用（2）式计算检验统计量为

$$T=\frac{(12-10.5)^2}{10.5}+\frac{(18-19.5)^2}{19.5}+\frac{(15-15.5)^2}{15.5}+\frac{(5-4.5)^2}{4.5}=0.401 \qquad (12)$$

临界域对应着 $T>\chi^2_{0.95}$（$c-1-k$）$=\chi^2_{0.95}$（$4-2-1$）$=\chi^2_{0.95}$（1）$=3.841$（自由度为 1 的 χ^2 分布的 0.95 分位数，由表 A2 获得），而 $T=0.401<3.841$，所以应接受 H_0，且 p-值大于 0.25.

用分组数据估计方差，当中间区间段等宽时（比如长度都为 h），这时通常用 Sheppard 修正法做一些修正. Sheppard 修正法是将 S^2 减去 $h^2/12$，以获得较好的方差估计. 本例中，$h=20$（区间宽度），所以在第二步可将方差减去 $20^2/12=33.33$，然后求平方根，结果为 $S=17.8$，是 σ 的一个较为小些的估计，这个偏小的 σ 的估计会使本例中的 T 增大. 因为我们的目标是获得使 T 尽可能小的估计，所以本例中我们没有用这个修正. 在很多情形下，我们希望修正后的 T 变得更小.

本例的另一个特点是用 $\overline{X}=55.04$ 和 $s=19.0$ 分别估计 μ 和 σ 时，可以获得 T 的一个较小值（0.279），55.04 和 19.0 是分组前由原始数据计算样本矩得到的，不管我们怎样得到它们，所用的这些估计是使得 T 变小的估计. 在很多情形中，本例中用到的步骤可以依赖于提供一个与 T 的最小值差异不大的 T 值. 所以这种处理步骤是值得推荐的. ■

上例中，当我们用分组前（不是分组后）原始数据得到的 $\overline{X}$ 和 S 来估计 μ 和 σ 时，检验统计量碰巧最小，所以上述处理步骤也曾被 Yule 和 Kendall（1950）推荐使
248 用，但是在使用分组数据时，其他一些用于分组数据的方法往往优于它，见（Chernoff 和 Lehmann 1954）.

□**理论** 如果零假设中假设的分布完全给定，若 H_0 为真，则分类确定后，观测在每类中的概率 p_j 也就已知. 对于样本容量为 N，每类中观测数分别为 O_1，O_2，…，O_c 的概率为：

$$P(O_1,O_2,\cdots,O_c|N)=\frac{N!}{O_1!O_2!\cdots O_c!}p_1^{O_1}p_2^{O_2}\cdots p_c^{O_c} \qquad (13)$$

这是一个多项分布，是将数据分成两类满足二项分布的情形，推广到将数据分成 c 类情形的分布. 由（13）式中的分布函数，可得 T 的概率分布，当 N 和 c 很大时，计算虽很复杂但 T 的概率分布仍可得到. 关于从样本中估计几个参数时，寻求 T 的精确分布似乎没有什么理论发展，因此，在用拟合优度检验方法时，大样本逼近既实用又必要. 关于大样本 χ^2 逼近理论可参考 Cramér(1946). □

Slakter(1966,1968), Dahiya(1971)和 Gurland(1972,1973), Pahl(1969)以及 Koehler 和 Larntz(1980)讨论了具有较小期望观测频数情形时的χ^2 拟合优度检验. 所有的 $E_j = 1$ 情形的精确表可查阅 Zahn 和 Roberts(1971). 如果样本数据是根据观测出现的时间而不是观测值分组时, 则检验的使用需要做一些修正(Putter,1964). Chernoff (1967)讨论了估计参数后自由度的调整. 关于χ^2 拟合优度检验的进一步探讨可参考 Efron 和 Morris(1975), Molinari(1977)以及 Hewett 和 Tsutakawa(1972), 同其他拟合优度检验的比较可参考 Holst(1972), Cohen 和 Sackrowitz(1975)及 Horn(1977).

习题

1. 检验下列数据看是否来自于 0.0000 和 0.9999 之间均匀分布的总体.

0.4755	0.5233	0.5440	0.5456	0.9056
0.2186	0.7500	0.2484	0.5101	0.8283
0.5112	0.5484	0.5758	0.3607	0.4352
0.3826	0.6454	0.9145	0.3943	0.5381
0.5758	0.8620	0.6687	0.3979	0.5646
0.4274	0.5482	0.3007	0.4438	0.4102
0.4295	0.5926	0.6521	0.6328	0.5689
0.7297	0.3768	0.8403	0.2925	0.2113
0.8757	0.4403	0.4993	0.3900	0.5166
0.8230	0.8522	0.8312	0.7979	0.4632
0.8432	0.4004	0.4295	0.9763	0.5590
0.4396	0.2595	0.3003	0.3003	0.5836
0.5337	0.8008	0.4887	0.2172	0.9329
0.5498	0.3686	0.4067	0.5274	0.4579
0.9096	0.4995	0.2172	0.6793	

249

2. 掷 600 次骰子, 得到如下结果:

出现的点数	1	2	3	4	5	6
频率	87	96	108	89	122	98

问这枚骰子是否均匀?

3. 用例 4.5.2 中每个运动员的击中数来检验如下零假设, H_0:18 个运动员击中的概率相同. 注意二项分布的一个假定是对所有基本试验来说概率相同, 这也是检验例 4.5.2 中假设的一种方法.

4. 没有书和表的帮助, 试着写 300 个随机个位数字, 然后用例 4.5.2 中使用的随机性检验来判断你是不是一个好的随机数生成器.

5. 去年在 Methodist 医院出生的婴儿数如下:

冬季 36 春季 45 夏季 42 秋季 55

检验如下假设, H_0:去年四个季度婴儿出生数服从均匀分布.

6. 现有 26 个观测的样本, 要检验它是否来自均值为 12, 标准差为 3 的正态分布总体, 已知没有一个观测在这个分布的下四分位数以下, 有 12 个观测在上四分位数以上, 6 个观测低于中位数, 8 个观测位于中位数与上四分位数之间. 问这些观测是来自于上述的正态分布吗?

4.6 相关观测的 Cochran 检验

有时，给对象某种处理或条件的使用会导致两个结果，例如对销售技巧有“卖”与“不卖”两种反应，或给对象某种处理可能导致“成功”或“失败”两个结果. 当然，在几个不同且独立的基本试验中，如果有 c 个处理，且每个处理得到两个结果，则整个结果可以用 $2\times c$ 列联表表示，其中一行表示成功数，另一行表示失败数. 并且可用4.2节所描述的 χ^2 列联表检验来检验零假设：无处理差异. 然而，我们经常要区分处理方法间更微小的差异. 这就需要对相同的区组独
250 立地用所有 c 种处理来提高检验的功效，比如在试验中用 c 种销售技巧向每位顾客推销产品，然后记录下每种技巧对每个人的影响. 因此每一块或每个人，在自愿状态下做出反应，这样的处理也就更有效. 这种试验性技巧称为“区组化”，我们称这种试验设计为“随机化完全区组设计”. 如果处理结果可归入两类中的某一类，下面的检验方法是一个合适的分析方法，这个检验方法称为 Cochran 检验，它是由 Cochran(1950)提出.

▶Cochran 检验

数据 独立地用 c 种处理方法分别处理 r 区组或 r 个对象，每种处理后的结果根据“成功”和“失败”（或其他可能处理结果的两种区分）分别记为“0”和“1”，然后将结果放入一张 $r\times c$ 列联表中，其中行代表每区组，c 列代表 c 种处理方式，每格的值要么为0，要么为1. 记 $R_i, i=1,2,\cdots,r$ 为行总和，$C_j, j=1,2,\cdots,c$ 为列总和，则数据可整理如下：

	处理				
区组	1	2		c	行总和
1	X_{11}	X_{12}	...	X_{1c}	R_1
2	X_{21}	X_{22}	...	X_{2c}	R_2
...	...	...	...	...	...
r	X_{r1}	X_{r2}	...	X_{rc}	R_r
列总和	C_1	C_2		C_c	$N=$总和

其中，$X_{ij}=0$ 或 1，N 为表中 1 的总数.

假定条件

1. r 个区组是从所有区组组成的总体中随机选取的.

2. 处理的结果可以按照某种方式对每个区组内的所有处理进行两种区分，所以结果可记为“0”或“1”.

检验统计量 检验统计量 T 可写为

$$T = c(c-1)\frac{\sum_{j=1}^{c}\left(C_j - \frac{N}{c}\right)^2}{\sum_{i=1}^{r} R_i\left(c - R_i\right)} \tag{1}$$ 251

下面的表达式用于计算更加合适

$$T = \frac{c(c-1)\sum_{j=1}^{c} C_j^2 - (c-1)N^2}{cN - \sum_{i=1}^{r} R_i^2} \tag{2}$$

零假设 T 的精确分布是难以求得的，所以我们用大样本逼近后的分布来代替，即假设区组数 r 相当大，则零分布可近似为自由度为 $c-1$ 的χ^2 分布.

假设

H_0:所有的处理效果相同

H_1:处理之间效果有差异

我们可以用数学语言来描述上述假设，记 $p_j = P$（列 j 中出现“1”的概率），则所有处理之间等效果的假设可描述为：

H_0:在每个区组中有 $p_1 = p_2 = \cdots = p_c$

处理间效果有差异即为：

H_1:对某两个处理 i 和 j 有 $p_i \neq p_j$

若 $T > \chi^2_{1-\alpha}(c-1)$（自由度为 $c-1$ 的χ^2 分布的 $1-\alpha$ 分位数，可由表 A2 获得），则拒绝 H_0，p-值近似为一个自由度为 $c-1$ 的χ^2 分布的随机变量大于 T 的概率，它可从表 A2 中查得.

多重比较 若拒绝了 H_0，则可用 McNemar 检验对各种处理方法进行成对比较，McNemar 检验是一种双边符号检验，在 3.5 节中可以查到.

计算机辅助 Cochran 检验可在 *StatXact* 中找到. ◀

例 4.6.1

3 个篮球爱好者分别设计了一个系统来预测学院篮球比赛的结果. 他们随机选取 12 场比赛并让每位运动员对每场比赛的结果作出预测，所有比赛结束后的结果记录如下（1 代表成功的预测，0 代表错误的预测）： 252

	运动员			
比赛	1	2	3	总和
1	1	1	1	3
2	1	1	1	3
3	0	1	0	1
4	1	1	0	2
5	0	0	0	0
6	1	1	1	3
7	1	1	1	3

8	1	1	0	2
9	0	0	1	1
10	0	1	0	1
11	1	1	1	3
12	1	1	1	3
总和	8	10	7	25

因为比赛（区组）是从正在进行的学院比赛中随机选取的，所以满足 Cochran 检验的假设条件. 因此我们用 Cochran 检验来检验零假设：

H_0:每个运动员以他的能力来预测篮球比赛是等有效的

用（1）式计算检验统计量为

$$T = c(c-1)\frac{\sum_{j=1}^{c}\left(C_j - \frac{N}{c}\right)^2}{\sum_{i=1}^{r} R_i(c - R_i)}$$

$$= \frac{(3)(2)[(-\frac{1}{3})^2 + (\frac{5}{3})^2 + (-\frac{4}{3})^2]}{2+2+2+2+2} = 2.8 \tag{3}$$

近似水平为 0.05 的临界域对应着 $T > \chi^2_{0.95}(2) = 5.99$（自由度为 2 的 χ^2 分布的 0.95 分位数，可由表 A2 获得），因为 $T = 2.8 < 5.99$，所以接受 H_0，因此，我们得出所用的预测方法没有显著差异，在 α 约为 0.25 时，H_0 也被拒绝，故 p-值为 0.25. ■

□**理论** 每个所定义的 X_{ij} 服从参数为 p 的伯努利分布（$n=1$ 的二项分布），在零假设下，每一行内所有的 X_{ij} 都是一样的，但是当区组变化时，X_{ij} 可以不同. 定义列总
253 和 C_j 为

$$C_j = \sum_{i=1}^{r} X_{ij} \tag{4}$$

因此 C_j 也是随机变量. 因为 C_j 是 r 个独立随机变量之和，所以当 r 较大时，利用中心极限定理得出 r 的分布近似为正态分布，即

$$\frac{C_j - E(C_j)}{\sqrt{\mathrm{Var}(C_j)}}$$

的分布函数近似为标准正态分布函数，并且，由定理 1.5.3 知，和式

$$\sum_{j=1}^{c}\left[\frac{C_j - E(C_j)}{\sqrt{\mathrm{Var}(C_j)}}\right]^2 = \sum_{j=1}^{c}\frac{[C_j - E(C_j)]^2}{\mathrm{Var}(C_j)} \tag{5}$$

可以用自由度为 c 的 χ^2 分布近似. 然而，参数 $E(C_j)$ 和 $\mathrm{Var}(C_j)$ 未知，所以下面的参数估计方法将导致减少 1 个自由度，这个方法由 Blomqvist(1951) 给出.

$E(C_j)$ 可由样本均值来估计

$$\frac{1}{c}\sum_{j=1}^{c} C_j = \frac{N}{c} = E(C_j) \text{ 的估计} \tag{6}$$

同样的估计用于估计每个 $E(C_j)$，$j = 1, 2, \cdots, c$. C_j 的方差为 X_{ij} 的方差对 j 求和，即

$$\mathrm{Var}\,(C_j)=\sum_{i=1}^{r}\mathrm{Var}\,(X_{ij}) \tag{7}$$

因为当区组变化时，X_{ij}相互独立（见定理 1.4.3），所以由（1.4.8）式给出的 X_{ij}的方差为

$$\mathrm{Var}\,(X_{ij})=p(1-p) \tag{8}$$

在零假设 H_0 下，每行中所有列“成功”的概率相同. 因此对每一行，我们自然地用行 i 成功的平均数 R_i/c 来估计 p，即

$$\text{行}\,i\,\text{中}\,p\,\text{的估计}=R_i/c \tag{9}$$

且由（8）式

$$\mathrm{Var}\,(X_{ij})\text{的估计}=\frac{R_i}{c}\left(1-\frac{R_i}{c}\right) \tag{10}$$ 254

然而，这种估计有偏小的趋势，所以我们将（10）乘上因子 $c/(c-1)$，则 $\mathrm{Var}(X_{ij})$ 由下式估计：

$$\mathrm{Var}\,(X_{ij})\,\text{的估计}=\frac{R_i\,(c-R_i)}{c(c-1)} \tag{11}$$

且由（7）式知，$\mathrm{Var}(C_j)$的估计可取为：

$$\mathrm{Var}\,(C_j)\text{的估计}\ =\frac{1}{c(c-1)}\sum_{i=1}^{r}R_i(c-R_i) \tag{12}$$

它与 j 无关，所以对所有 C_j 都成立，将 $E(C_j)$（(6)式）和 $\mathrm{Var}(C_j)$（(12)式）的估计带入（5）式中，得到

$$\begin{aligned} T&=\sum_{i=1}^{c}\frac{\left(C_j-\dfrac{N}{c}\right)^2}{\displaystyle\sum_{i=1}^{r}\frac{R_i(c-R_i)}{c(c-1)}}\\ &=c(c-1)\frac{\displaystyle\sum_{j=1}^{c}\left(C_j-\frac{N}{c}\right)^2}{\displaystyle\sum_{i=1}^{r}R_i(c-R_i)} \end{aligned} \tag{13}$$

这就提供了统计量 T 的分布可由自由度为 $c-1$ 的χ^2 分布近似的某种直观理解. □

Berger 和 Gold(1973)及 Bhapkar 和 Somes(1977)探讨了 Cochran 检验，Patel(1975)讨论了检验统计量的精确分布. Cochran 检验方法用于其他模型的例子可参考 Fleiss(1965). Tate 和 Brown 考虑了大样本逼近问题.

评注

若仅考虑两种处理，比如同一区组（r 区组）上的观测有“处理前”与“处理后”两种情况，试验情形也与 McNemar 对变化的显著性检验中所分析的相同，即每种情况下零假设是：总体在类 1 中的比例在处理 1（处理前）和处理 2（处理后）相同. 因此如果 $c=2$，试验者可选用 Cochran 检验或 McNemar 检验. 事实上，

当$c=2$时，则没有选择，此时 Cochran 检验与 McNemar 检验（见 3.5 节）相同，
255 原因如下.

当 $c=2$ 时，Cochran 检验统计量变为

$$T=2\frac{\left(C_1-\frac{C_1+C_2}{2}\right)^2+\left(C_2-\frac{C_1+C_2}{2}\right)^2}{\sum_{i=1}^{r}R_i(2-R_i)}$$

$$=2\frac{\left(\frac{C_1}{2}-\frac{C_2}{2}\right)^2+\left(\frac{C_2}{2}-\frac{C_1}{2}\right)^2}{\sum_{i=1}^{r}R_i(2-R_i)}$$

$$=\frac{(C_1-C_2)^2}{\sum_{i=1}^{r}R_i(2-R_i)} \tag{14}$$

若某区组在两列中都为1，则 $R_i=2$ 且 $R_i(2-R_i)=0$. 同理，若两列都为0，则 $R_i=0$ 且 $R_i(2-R_i)=0$. 若在某行中两列分别为0，1或1，0，则 $R_i=1$ 且 $R_i(2-R_i)=1$. 因此（14）式的分母只包含两列分别为0，1或1，0的行总和，即 McNemar 检验中的 $b+c$（记号），且 C_1 是第一列中或“处理前”的个数，即 McNemar 检验中的 $c+d$（记号），类似地，$C_2=b+d$，因此我们有

$$C_1-C_2=c+d-b-d$$

且（14）式可化为

$$T=\frac{(c-b)^2}{b+c}=\frac{(b-c)^2}{b+c}$$

这与（3.5.1）式给出的 McNemar 检验统计量形式相同，当 $c=2$ 时，McNemar 检验统计量与 Cochran 检验统计量都可由自由度为1的 χ^2 分布近似.

习题

1. 同时对12个家庭主妇自愿者使用两种不同的销售技巧，要检验这两种技巧的相对有效性. 两种销售技巧目的是让每个家庭主妇买某种产品，在整个过程中，产品不变，试验结束后，如果某个技巧令一个家庭主妇同意购买这种产品，则记为1，否则记为0，结果
256 如下：

家庭主妇	1	2	3	4	5	6	7	8	9	10	11	12
技巧1	1	1	1	1	1	0	0	0	1	1	0	1
技巧2	0	1	1	0	0	0	0	0	1	0	0	1

（a）用 Cochran 检验.

（b）重新安排数据，然后使用（3.5.1）式建议的大样本形式的 McNemar 检验.

（c）忽略这个试验中的区组效应，将数据看成使用24个不同家庭主妇的结果，用第4.1

节中的概率差异检验来分析数据，与 Cochran 检验相比较并做讨论.

2. 在一条船上，将水手随机地分为 12 组，每组 3 人，每组的水手都在船的同一地方干相似的工作. 对每组的每个水手随机地使用治疗方法 1，2 或 3，且对同组中三个水手的治疗两两不同，治疗方法 1 是“打针”，治疗方法 2 是“吃药”，治疗方法 3 是“没有患感冒者有两个星期的假期”. 只要每个水手报告他患感冒的情况，试验者就要做一份治疗报告. 冬天结束后，结果如下表：

组	患感冒的水手（治疗次数）
1	2
2	1, 2
3	1, 2, 3
4	2, 3
5	2
6	无
7	1, 2
8	1, 2
9	1
10	2
11	1, 2, 3
12	2

这些结果表明不同治疗手段间有显著差异吗？

3. 用计算机生成 100 组人工数据以比较 3 种统计检验方法的相对功效. 对每组数据分别用这 3 种检验（置信性水平 α 为 0.05），是否接受零假设的记录如下：

检验 1	检验 2	检验 3	数据组数	
接受	接受	接受	26	
接受	接受	拒绝	6	
接受	拒绝	接受	12	
拒绝	接受	接受	4	257
拒绝	拒绝	接受	18	
拒绝	接受	拒绝	5	
接受	拒绝	拒绝	7	
拒绝	拒绝	拒绝	22	

问这 3 种检验用于所得的模拟数据的总体时，其检验的功效是否有显著差异？

思考题

1. 在每个处理区组的组合中，替用一个观测，现在在每个格子中，我们有 m 个独立的观测，令 C_j，R_i 和 N 分别代表行处理方法总数，行总和与所有和，这与以前定义的一样. 证明统计量

$$T' = mc(c-1)\frac{\sum_{j=1}^{c}\left(C_j - \frac{N}{c}\right)^2}{\sum_{i=1}^{r} R_i(mc - R_i)}$$

同 T 的分布一样也可由自由度为 $c-1$ 的 χ^2 分布近似.

2. 对本节所设计的常用参数检验，一般假设观测样本来自于正态分布总体，而不是 $n=1$ 的二项分布总体，并使用"F-统计量"检验. 如果观测为0或1，则F-统计量可简化为

$$F=(r-1)\frac{c\sum_{j=1}^{c}C_j^2-N^2}{rcN-r\sum_{i=1}^{r}R_i^2-c\sum_{j=1}^{c}C_j^2+N^2}$$

证明 F 与 T 有如下函数关系

$$F=\frac{(r-1)T}{r(c-1)-T}$$

并且对太大的 T 拒绝 H_0 等价于对太大的 F 拒绝 H_0.

4.7　其他分析方法讨论

似然比统计量

258 本章所描述的方法不限于分析列联表，概括起来说，所使用的检验统计量为：

$$T_1=\sum_{\substack{\text{所有}\\\text{格子}}}\frac{(O_i-E_i)^2}{E_i} \tag{1}$$

这里 O_i 为格子 i 中的观测数，E_i 为格子 i 的期望观测数. 检验统计量 T_1 由 Pearson (1900,1922)引入，为了与其他 χ^2 统计量区分而被称为"Pearson χ^2 统计量". 下面我们将介绍一种其他的 χ^2 统计量.

有一种不同的分析方法，我们称为似然比检验法（在问题4.2.3中提到过），用统计量

$$T_2=2\sum_{\substack{\text{所有}\\\text{格子}}}O_i\ln\left(\frac{O_i}{E_i}\right) \tag{2}$$

代替 T_1，这里"ln"代表自然对数，很多计算器可计算. 统计量 T_2 同 T_1 一样渐近地服从同样自由度的 χ^2 分布，尽管两个统计量有相同的渐近分布，但对某一特定的列联表它们的值可能有很大不同. 选择使用统计量 T_1 和 T_2 主要看使用者的个人喜好.

统计量 T_2 也称为"似然比 χ^2 统计量"，这是因为它来源于统计学中的似然比理论. 它属于 Wilks(1935,1938)，并因为源于似然比理论而受到广泛使用. 然而，使用统计量 T_2 的一个严重的弊端就是，如果 $N/rc<5$，χ^2 分布近似效果就会不好，但是对统计量 T_1 来说，即使 N 值偏小，χ^2 分布近似效果仍然不错. Agresti(1990)说明了如果 r，c 太大使 $N/rc\leqslant 1$，且"列联表中既不包含太小也不包含适当大的期望频

数”时，那么 T_1 的分布用 χ^2 分布近似的效果不好. Agresti 的发现同我们建议的所有期望值至少为 0.5 且大多数大于 1.0 不一致.

对数线性模型

还有一个广泛使用的分析方法是“线性对数模型”. 如果有适当的计算机程序帮助计算，这种方法能很好地分析三维以上的列联表，我们不推荐用手工计算. 也可在线性对数模型中使用前面介绍的统计量 T_1 和 T_2；不同之处在于得到所有 E_i 所使用的方法. 通常我们使用迭代法，因此这需要计算机的帮助. *SAS*, *StatMost* 和 *SYSTAT* 都有计算机程序来完成多维列联表的对数线性分析. 259

对数线性模型的名字来自于下面的原因，在一个双向列联表中，独立性的零假设可以用数学语言表示如下：

$$H_0: p_{ij} = p_{i+} \cdot p_{+j}, \text{对所有的 } i \text{ 和 } j,$$

这里 p_{ij} 是观测归入格（i,j）中的概率，p_{i+} 和 p_{i-} 分别为行列边际概率. 对零假设两边取对数得

$$H_0: \log p_{ij} = \log p_{i+} + \log p_{+j}$$

这是一个线性等式，则对零假设的检验归结为检验格子概率的对数是否是边际概率的线性函数. 关于对数线性模型的完整描述和分析可参考 Bishop, Fienberg 和 Holland (1975). 建议初学者可参考 Ku 和 Kullback(1974) 或 Lee(1978) 对问题的初等处理，也可参考 Fienberg(1977) 所写的书.

感兴趣的读者想进一步了解对数线性模型，可阅读下面的文章：Bishop(1969, 1971)，Fienberg(1970b, 1972)，Fienberg 和 Larntz(1976)，Chen 和 Fienberg(1976)，Grizzle, Starmer 和 Koch(1969)，Koch 和 Reinfurt(1971)，Grizzle 和 Williams(1972)，Koch, Imrey 及 Reinfurt(1972)，Wagner(1970)，Odoroff(1970)，Goodman(1971)，Gart(1972)，Haberman(1973) 和 Read(1977)，以及 Agresti(1990) 的书.

4.8　第3、4章复习题

1. 丹麦一项研究计划试图找出酗酒是否与遗传因素有关. 5 位精神病医师专门对那些从婴儿时期就与亲生父母分开的男子做调查研究，一组中有 55 名男子的父亲或母亲有酗酒的癖好，而这 55 名男子中有 10 名酗酒. 而另外一组中有 78 名男子的父母都不是酗酒者，但 78 名男子中有 4 名酗酒. 研究发现第一组中酗酒有遗传是显著的. 你的统计分析结果如何？（引自美联社 1973 年 2 月 21 日的报导.）
2. 大选前随机地抽取 200 个选举人的样本，询问这 200 个人愿意投哪个候选人的票. 其中 85 人投候选人 A，111 人投候选人 B，4 票弃权. 你如何预测选举结果？（讨论需要讨论的内容.）
3. 向某个社区学院的几个非大一的学生提了几个问题，包括他们对把吸食大麻合法化有什么感想，结果如下 260

学生数	性别	从家到社区学院的距离	喜欢的政党	大麻问题	大一GPA
1	M	32	N	1	2.66
2	M	10	D	1	3.18
3	M	28.5	R	1	2.15
4	M	3.5	R	2	1.61
5	M	4	D	4	1.54
6	F	7	N	5	2.12
7	M	3.5	R	3	1.35
8	M	10	N	4	2.26
9	F	6	R	4	2.70
10	M	32	D	4	2.84
11	M	22.5	D	4	2.60
12	M	7	D	1	1.13
13	M	6.5	N	4	0.81
14	M	5	D	1	3.11
15	M	35	R	1	2.47
16	M	5.5	D	5	3.15
17	M	26.5	D	4	2.33
18	F	24	N	5	2.46
19	M	32	D	5	3.59
20	F	5	R	1	2.00
21	M	5.5	R	1	2.90
22	M	11.5	N	5	3.26
23	M	9.5	R	4	2.71
24	M	25.5	O	3	2.22
25	M	15	R	1	3.00
26	F	9	N	4	2.06
27	M	15	D	1	1.75
28	M	24	R	3	2.42

1 = 强烈反对，. . . ，5 = 完全同意.

(a) 检验假设：政党偏好与对大麻合法化的态度独立.

(b) GPA 与社区距离有关吗？

(c) 估计社区距离的中位数.

(d) 估计女学生的百分比.

(e) 政党偏好与性别（男或女）独立吗？

261 (f) 检验假设：男女学生有相同的大一学生 GPA.

4. 10 名学生需要上一门必修课程，每位学生都有一个平时成绩 X 和考试成绩 Y，记录如下：

	学生数									
	1	2	3	4	5	6	7	8	9	10
X	92	80	74	85	71	68	81	82	80	81
Y	94	86	72	91	70	77	89	91	86	94

(a) 考试成绩是否显著高于平时成绩?

(b) 记 $p=$ 一个学生考试成绩高于平时成绩的概率，求出 p 的一个置信区间.

(c) 找出平时成绩中位数的一个置信区间.

(d) Y 的上四分位数是否显著高于75?

5. 考虑如下医学试验：通过解剖动物尸体，测量它们的肝脏铁的吸收量，来决定乙硫氨基酪酸是否对动物的饮食有影响. 随机选取34只动物将其平均分为两组，其中17只动物的食物中含有乙硫氨基酪酸，另外17只没有. 将动物分别配对（一组的1对应另一组的1等），并给每对喂等量的食物. 一段时期后，取出动物的肝脏并用一个温度为"温"（37度）或"凉"（25度）的含放射性铁溶液进行处理. 不同肝脏铁的吸收量数据如下：

温			凉		
配对	含乙硫氨基酪酸	不含	配对	含乙硫氨基酪酸	不含
1	2.59	1.40	9	6.77	4.71
2	1.54	1.51	10	4.97	1.60
3	3.68	2.49	11	1.46	0.67
4	1.96	1.74	12	0.96	0.71
5	2.94	1.59	13	5.59	5.21
6	1.61	1.36	14	9.56	5.12
7	1.23	3.00	15	1.08	0.95
8	6.96	4.81	16	1.58	1.56
			17	8.09	1.68

(a) 在"温"的溶液中吃含乙硫氨基酪酸食物的动物的肝脏是否比不吃的那组更容易吸收溶液中的铁? 在"凉"的溶液中情况如何?

(b) 温度低的溶液是否能显著提高吃含乙硫氨基酪酸食物的动物的肝脏对铁的吸收? 对另一组中的动物情况又如何?

(c) 求吃含乙硫氨基酪酸食物的动物的肝脏被含放射性铁的溶液处理后，含铁量的一个双边容忍限. 262

(d) 同对中的两只动物肝脏的铁吸收量是否相关?

6. 科学家想了解爱斯基摩人的住房内的空气质量，于是他们做了一个试验，考察了20个位于Bethel（阿拉斯加州西南部Kuskokwima河边的一个土著村）的爱斯基摩人居住的房屋. 其中10个是作为房屋发展项目的一部分而新建的房子，其他10间是标准的Bethel地区土著住房. 试验目标是比较新旧住房每立方英尺空气中的细菌聚居数，看它们是否存在差异. 数据如下：

老房编号	每立方英尺细菌聚居数	新房编号	每立方英尺细菌聚居数
1	37.0	1N	1.0
2	2.6	2N	5.3
3	48.6	3N	3.4
4	47.8	4N	2.3
5	99.3	5N	5.1
6	1.4	6N	38.7
7	2.3	7N	5.0
8	3.1	8N	50.6
9	3.0	9N	1.6
10	0.3	10N	22.7

(a) 分析数据.(b) 求老房细菌数中位数的一个置信区间.

7. G. Noether 博士建议用如下趋势性检验，将观测结果分到互不重叠的，由3个相邻数组成的多个组中. 令 T 等于单调组（单调增或单调减）数，比如：

$$\underbrace{42,44,63,}_{\text{增加}}\ \underbrace{61,44,52,}_{\cdot}\ \underbrace{73,72,46,}_{\text{减少}}\ \underbrace{48,42,53}_{\cdot}$$

这里 $T=2$.

(a) 在随机变量是独立同分布的零假设下，T 的分布是什么？

(b) 如何找临界域？

8. 从工商管理学院拥有的计算机中随机抽取12台，其中有8台是IBM型号. 从医学院随机抽取的36台计算机中有30台是IBM型号. 问：

(a) 两个学院拥有IBM型号计算机的差异是否显著？

(b) 求医学院拥有IBM计算机的整体比例的一个置信区间.

9. 体育系将投掷标枪加入到课程中，并决定在运动场上应标出多少根标线来标出投掷距离.
263 他们要随机地挑选几名学生试投，并在学生投掷的最远和最近处标上标线. 为了能以90%的把握保证学生投掷标枪的距离至少有95%在标线所在范围内，问应选多少学生？

10. 一名经纪人记录了两年内他每月售出的地方性债券数如下：

	一月	二月	三月	四月	五月	六月
1997	12	16	14	18	18	14
1998	19	22	20	17	18	20

	七月	八月	九月	十月	十一月	十二月
1997	10	21	12	18	17	17
1998	20	16	16	21	24	25

这个记录是否表明他售出的债券数呈递增趋势？

11. 科学家做了一个检验大猩猩是否具有识别字母能力的试验. 他们将5个不同的字母随机地放在5个按钮上，当大猩猩按在字母E所在的按钮上时，小灯亮起来，然后猩猩会得到一个它喜欢的香蕉. 每天做5次试验，并且字母在每次试验后都会被随机排列一次. 试验持续了6天，结果如下：

正确选择E之前按按钮的次数

试验次数	1	2	3	4	5
星期一	6	4	4	2	3
星期二	7	8	6	1	3
星期三	4	2	1	2	3
星期四	1	4	3	2	2
星期五	5	2	3	1	2
星期六	4	2	1	2	1

(a) 大猩猩在每天的5次试验中识别字母的能力是否有所提高？

(b) 大猩猩在一周内识别字母的能力是否有所提高？

(c) 如果大猩猩随机地按按钮，则按按钮的次数服从几何分布，$P(X=k)=(0.2)(0.8)^{k-1}, k=1,2,3\cdots$. 检验假设：大猩猩是随机地按按钮.

12. 一条流水线上4套检验产品瑕疵的系统正在同时运行，每个产品都会被4套系统分别检验. 假设系统没有第Ⅰ类错误，即没有瑕疵的产品不会被系统误检，但是有第Ⅱ类错误，即瑕疵可能没有被检测到. 我们抽取了一组数据如下： 264

	系统1	系统2	系统3	系统4
产品1	瑕疵		瑕疵	瑕疵
产品2			瑕疵	
产品3				
产品4	瑕疵	瑕疵	瑕疵	
产品5	瑕疵			瑕疵
产品6	瑕疵			
产品7				
产品8	瑕疵		瑕疵	

问：4个系统发现瑕疵产品的能力是否有差异？

13. 银行从一台自动提款机上随机地抽取顾客的存款信息，并希望找出上四分位数的90%置信区间. 存款金额如下：

748	320	45	1,237	5,883	170
65	30	83	186	598	8,500
1,500	4,857	300	100	395	2,450
349	50	25	637	600	260
67	200	45	400	57	580

求出希望的置信限.

14. 对问题13中的随机样本，检验假设：用户存款金额低于100美元的概率等于存款超过1000美元的概率. 单边备择假设：用户存款金额低于100美元的概率高于存款超过1000美元的概率.

15. 由50名男孩与他们的父亲组成的随机样本，其中有17名男孩吸烟，27人的父亲吸烟，17名吸烟的男孩中有12位其父亲也吸烟. 男孩的吸烟习惯与父亲的吸烟习惯是否显著（正）相关？这是哪种类型的列联表？

16. 第50街必胜客最近雇了12个男孩和8个女孩，他们需要6名司机和14名收银员，他们安排6名男孩当司机. 检验零假设：工作安排与性别独立. 单边备择假设为：男孩更有可能被安排为司机. 求出p-值.

17. 轮盘赌的游戏中，得到红色的概率为18/38，得到黑色的概率也为18/38，得到绿色的概率为2/38. 500次转动的结果为：35次出现绿色，241次出现红色，其余为黑色. 问：观测概率是否与理论概率一致？

18. 假设环保局要求你们公司生产的汽车至少有95%的汽车每加仑油能跑20英里以上. 你检验60辆汽车发现它们全都达到标准（每加仑能跑20英里以上），你能否以95%的把握保证你公司至少95%的汽车每加仑油能跑20英里以上.
 (a) 在置信水平$\alpha=0.05$下用一种假设检验. 265
 (b) 对同样的问题求出单边95%置信区间（如果你不知道怎样求单边置信区间，你可以求出双边90%置信区间，然后去掉另一端）.

19. 1997年间，Joe的餐馆中的顾客使用万事达信用卡14次，维萨卡10次，发现卡4次及

美国运通卡1次. 1998年间，顾客使用万事达卡22次，维萨卡23次，发现卡10次，没有人使用美国运通卡. 从1997年到1998年信用卡的使用方式是否有显著的变化？你所使用检验的名字是什么？

20. 从某所大学即将入学的大一学生中随机选取140名新生，其中80名学生驾私家车到校，其余没有. 这些新生中有40名考试不合格，其余的全部通过. 这些数据的Phi系数为0.32. 开私家车到校与考试合格间是否有显著的关系？用双边检验.

21. 选取20封邮件用来测试写上邮政编码的邮件是否比没有邮政编码的邮件更早到达目的地. 将信配成10对，给每对邮件写上相同的地址并在同一时间邮寄，但一封邮件有邮编，另一封邮件没有，并将它们发到全美10大城市去，结果如下：

	有邮政编码	没有邮政编码
亚特兰大	3天	4天
巴尔的摩	3天	4天
芝加哥	4天	4天
底特律	4天	5天
艾尔金	3天	5天
费城	5天	4天
加里	3天	3天
哈特福德	5天	7天
印第安纳波利斯	4天	5天
洛杉矶	3天	4天

(a) 用一种假设检验方法检验：有邮编的邮件比没有邮编的邮件更早到达目的地.

(b) 求含有邮编的邮件比没有邮编的邮件早到概率的95%置信区间.

(c) 检验假设：含有邮编的邮件到达目的地的时间的上四分位数是3天，单边备择假设为：含有邮编的邮件到达目的地的时间的上四分位数超过3天.

22. 从国内和国外申请管理学博士学位的学生总体中分别随机选取8名学生，用来检验国外学生是否比国内的学生GMAT分析考试的分数高. 他们的分数为（百分制）

266

国外	79,	86,	93,	96,	97,	97,	99,	99
国内	76,	80,	87,	88,	89,	92,	94,	96

用第3或第4章中的方法分析这些数据，用通常的技巧给出近似p-值，同时求出精确的p-值，并比较这两个值.

23. 40名学生参加一个标准测试，获得如下分数（假设把这群学生看作随机样本）

100	92	90	83	75	69	64	47
100	92	89	81	73	68	60	44
97	91	88	80	71	65	58	40
92	91	85	78	70	65	56	38
92	90	83	75	70	64	49	36

(a) 以往的上四分位数为85，但是老师怀疑这次成绩的上四分位数高于85，进行合适的假设检验.

(b) 求出上四分位数的95%置信区间.

(c) 检验零假设：分数为90分以上的概率等于40分以下的概率；单边备择假设为：分

数为 90 分以上的概率大于 40 分以下的概率.

24. 一个心理学家试图证明丈夫比妻子更喜欢对方母亲的陪同，而不是自己母亲. 选取一些夫妇作为随机样本，数据如下：

夫 妇	丈夫的喜好	妻子的喜好
Adams	对方母亲	对方母亲
Baker	对方母亲	母亲
Chase	对方母亲	母亲
Dodge	母亲	母亲
Evans	对方母亲	母亲
Forrester	对方母亲	对方母亲
Graves	母亲	母亲
Holland	母亲	母亲
Islip	对方母亲	对方母亲
Jacobs	对方母亲	母亲
Kraft	对方母亲	母亲
Lewis	对方母亲	母亲
Morris	母亲	母亲
Noonan	对方母亲	母亲
O'Neil	对方母亲	母亲
Procter	母亲	母亲
Quincy	母亲	母亲
Reed	对方母亲	对方母亲
Smith	母亲	对方母亲
Tracy	母亲	母亲
Unseld	对方母亲	母亲
Victor	对方母亲	母亲
Williams	母亲	母亲
Young	对方母亲	母亲
Zyskind	对方母亲	母亲

267

请使用合适的假设检验来分析这个数据.

25. 某便利店的经理相信外州的顾客（从车牌号能看出）更喜欢使用信用卡购买物品，他记录下一个下午所有顾客的付账方式，结果如下：

	卡	现金
州外	14	10
州内	6	18

经理的想法正确吗？这是哪种类型的列联表？ 268

第5章 秩 检 验

导 言

前几章中介绍的大部分统计方法都可以用于有名义尺度的数据. 第3章提出了几种分析自然对分数据的方法，即这些数据是0-1型或成功—失败型数据. 第4章的讨论集中在分析根据两个或更多不同标准分类的数据，且根据每个标准数据分成两个或更多的类. 所有这些方法可以用于信息多于可利用的名义型信息的数据，但是，由于各种原因，诸如要求计算的速度和方便，数据比较冗余，或者要求对数据有特殊的解释等，人们忽视了数据的一些信息，从而将数据简化为名义型数据来分析. 这种信息的损失经常导致相应功效的损失. 本章给出了几种统计方法，如果数据至少具有次序尺度，则这些方法会利用包含在数据中的更多信息.

数据可能是非数值的（“好，更好，最好”）或数值的（7.36,4.91，等等）. 如果数据是非数值的，但是依次序型数据排序，那么，本章的方法通常是有用方法中最有功效的. 如果数据是数值型的，而且是满足通常参数检验所有假定条件的正态
269 分布随机变量的观测值，用本章的方法所导致的效率损失将会非常小. 在那些情形下，只用观测值的秩做检验的相对效率经常是0.95，这要视情形而定.

本章的秩检验对各种类型的总体都有效，不论是连续型的，离散型的，还是二者的混合. 早期非参数统计中的结果为了使基于秩的检验有效，要求假设变量是连续型的. Conover(1973a)和其他人的研究结果表明，连续性的假设是没必要的，可以由一个很简单的假设 $P(X=x)<1, \forall x$ 来代替. 因为任何的试验者都不愿意从完全由单独一个数字构成的总体中抽样，所以在本章的检验中我们不会列出这个假设.

如果数据是有序的并且有很多结（如果两个观测值相等，则称它们为数据的结），则可以用秩检验来分析. 这里需要小心的是，只有当数据中没有结时，所谓的小样本“精确表”才是精确的，否则它们是近似的. 对给定结集的精确表可以按照没有结的方法得到，但是这种随结集而变化的一系列精确表并不实用. 如果数据中有大量的结，则本书运用大样本近似，而不用小样本表.

按升序排列观测值的一个很方便的方法是Tukey(1977)给出的茎叶图方法，或许最简单的解释茎叶图方法的方式就是给出一个例子. 假设一个班级28个学生在一次考试中所得的分数如下：

74	63	88	69	81	91	75
82	91	87	77	86	86	87
96	84	93	73	74	93	78
70	84	90	97	79	89	93

在这个例子中，每个考试分数的十位数是茎，有 4 个不同的茎：6,7,8 和 9，个位数是叶. 首先，主茎如下：

9
8
7
6 270

然后，每个叶写到对应茎的右边. 也就是将第一个分数 74 写成 4 靠在茎 7 的右边. 每个分数都用这种方式来写，结果如下：

9	6	1	3	0	7	1	3	3		
8	2	4	4	8	7	1	6	6	9	7
7	4	0	7	3	4	9	5	8		
6	3	9								

考试成绩的分布图立刻呈现出来. 但是，最重要的目的是现在可以很容易的从小到大排列分数，可以根据这种方式给观测值赋予秩. 这种茎叶图的方法可以使得本章的方法变得更容易使用.

5.1 两个独立样本

本节提出的检验是人们所熟知的 Mann-Whitney 检验和 Wilcoxon 检验. 可能部分是由于检验方法的直观性的要求，几乎在同一时期，该检验的许多等价形式以各种各样的称呼在文献中出现.

通常两样本情形是试验者从可能不同的总体中得到的两个样本，希望用一个统计检验来判断两个总体相同的零假设是否被拒绝. 也就是试验者希望以两个总体的随机样本为基础检验出两个总体间的差异. 一个等价情形是，得到一个随机样本，但它被随机的分为两个样本，一个样本得到一种处理，另一个样本得到不同的处理，例如，在一个医学试验中，对一组病人用新的药物治疗，对另一组病人用现有标准的药物治疗或根本不用药物. 分析这种类型的两样本试验和第一种类型的一样.

如果样本是由有序型数据组成的，则对两样本的差异最感兴趣的是两个总体位置的差异. 如一个总体的取值趋向于比另一个总体的取值趋向大吗？两个中位数相等吗？两个均值相等吗？

对于两样本问题，一个直观的方法是将两个样本组合成一个有序的样本，不管每个值是来自哪个总体，从小到大给样本值赋秩. 这样，检验统计量就可能是那些
来自其中一个总体观测值的秩和. 如果这个和太小（或者太大），就意味着来自这个 271
总体的值趋向于比来自另一总体的值小（或者大，根据具体情况而定）. 因此，如果与一个样本相关的秩趋向于比与另一个样本相关的秩大，则两个总体没有差异的零假设可能被拒绝.

对于实际数据人们更喜欢考虑秩，可能有如下原因. 首先，如果赋给观测值的数字本身没有意义，但只是与其他观测进行有序比较后才有意义，那么这些数字没

有秩所含的信息多；有序的数据自然也一样．第二，即使这些数字有意义而分布函数并不是正态分布函数，当检验统计量是基于实际数据时，概率理论通常也超出了我们能达到的范畴．基于秩的统计量的概率理论相对比较简单，而且在很多情形下并不依赖于分布．更喜欢秩的第三个原因是，与通常参数检验类似的两样本 t 检验相比时，Mann-Whitney 检验的渐近相对效率（A. R. E.）从来不会太坏．但反过来并不成立；与 Mann-Whitney 检验相比，t 检验的 A. R. E. 可能和0一样小，或者“无限的坏”．所以用 Mann-Whitney 检验更安全.

▶Mann-Whitney 检验

数据　数据由两个随机样本组成，记 $X_1, X_2, \cdots, X_n$ 为来自总体1、容量为 n 的随机样本，记 $Y_1, Y_2, \cdots, Y_n$ 为来自总体2、容量为 m 的随机样本．给这 $n+m$ 个观测从小到大赋予秩，记 $R(X_i)$ 和 $R(Y_j)$ 分别为赋给 X_i 和 $Y_j(\forall i,j)$ 的秩．为方便起见，令 $N=n+m$.

如果几个样本值完全相等（结），则给每个值赋予秩是在没有结时它们该有秩的平均（见例5.1.1）.

假定条件

1. 两个样本都是来自于各自总体的随机样本
2. 除了每个样本内观测相互独立外，两个样本之间也相互独立
3. 度量尺度至少是须序的

检验统计量　如果没有结，或者结很少，赋给来自总体1的样本秩和可以用作检验统计量.

272
$$T=\sum_{i=1}^{n} R(X_i) \tag{1}$$

如果有很多结，用 T 减去均值再除以标准差就得到

$$T_1=\frac{T-n\dfrac{N+1}{2}}{\sqrt{\dfrac{nm}{N(N-1)}\sum_{i=1}^{N} R_i^2-\dfrac{nm(N+1)^2}{4(N-1)}}} \tag{2}$$

其中，$\sum R_i^2$ 是所有在两个样本中实际用到的 N 个秩（或平均秩）的平方和.

零分布　当 $n \leqslant 20$，$m \leqslant 20$ 时，所选的 T 的零分布的下分位数在表 A7 中给出．T 的上分位数 w_p 由下面关系式得到

$$w_p=n(n+m+1)-w_{1-p} \tag{3}$$

此处，下分位数 w_{1-p} 由表 A7 获得.

作为用上分位数的备择，如下定义统计量 T'：

$$T'=n(N+1)-T \tag{4}$$

如果需要右边检验，它就可以和下分位数一起运用．用 T' 比从（3）式求上分位数和右边 p-值更方便.

只有当数据中没有结时，表 A7 中的分位数才是精确的，从而没有用到平均秩.

在没有结且 n 或 m 大于 20 的情况下，近似的分位数可由正态逼近得到，

$$w_p \cong \frac{n(N+1)}{2} + z_p\sqrt{\frac{nm(N+1)}{12}} \tag{5}$$

其中，分位数 z_p 可由表 A1 获得.

如果数据中有很多结，就用 T_1 代替 T，T_1 是近似服从标准正态的随机变量，它的分位数由表 A1 给出.

假设

A.（双边检验）

令 $F(x)$ 和 $G(x)$ 分别为 X 和 Y 的分布函数，则假设可表述如下：

$$H_0: F(x) = G(x) \qquad 对于任意\ x$$
$$H_1: F(x) \neq G(x) \qquad 对某些\ x$$

273

检验对 $H_1: E(X) \neq E(Y)$ 是敏感的，而且可以用作检验均值. 在很多实际情况中，分布间的差异表明 $P(X>Y)$ 不再等于 1/2. 因此经常用 $H_1: P(X>Y) \neq P(X<Y)$ 代替上边的假设.

如果 T 小于它的 $\alpha/2$ 分位数或大于它的 $1-\alpha/2$ 分位数，则以显著性水平 α 拒绝 $H_0, n \leqslant 20, m \leqslant 20$，时，分位数可以从表 A7 和（3）式得到，对于大样本逼近，分位数可以从表 A1 和（5）式得到. 如果用 T_1 代替 T，则分位数可以直接从表 A1 得到.

近似双边 p-值可以在表 A1 中找到. 对于 T，将 T 或 T' 中较小者用在下式中

$$p\text{-值} = 2 \cdot P\left(Z \leqslant \frac{T\,(\text{或}\ T') + \frac{1}{2} - n\frac{N+1}{2}}{\sqrt{\frac{nm(N+1)}{12}}}\right) \tag{6}$$

其中，Z 是标准正态随机变量. 对于 T_1，p-值是 2 倍的 $P(Z \leqslant T_1)$ 或 $P(Z \geqslant T_1)$ 中较小者.

B.（左边检验）

零假设是：

$$H_0: F(x) = G(x)$$

而左边检验的备择假设可以用下边形式中的一个

$$H_1: F(x) > G(x)$$
$$H_1: E(X) < E(Y)$$

或

$$H_1: P(X > Y) < P(X < Y)$$

全部都以不同的方式传达这样的思想，“X 趋向于比 Y 小”.

当 $n \leqslant 20$，$m \leqslant 20$ 时，如果 T 小于表 A7 给出的 α 分位数，则拒绝 H_0；当样本量更大时，如果 T 小于（5）式给出的 α 分位数，则拒绝 H_0. 如果用 T_1，则当 $T_1 < z_\alpha$ 时，以水平 α 拒绝 H_0，这里 z_α 从表 A1 得到.

p-值可由下面的概率近似：

274

$$p\text{-值} \cong P\left(Z \leqslant \frac{T + \frac{1}{2} - n\frac{N+1}{2}}{\sqrt{\frac{nm(N+1)}{12}}}\right) \tag{7}$$

这可从表 A1 得到. 对于 T_1，p-值近似于 P（$Z \leqslant T_1$），可直接从表 A1 得到.

C. （右边检验）

零假设是：

$$H_0: F(x) = G(x)$$

右边检验的备择假设可以用下面形式中的一个

$$H_1: F(x) < G(x)$$

$$H_1: E(X) > E(Y)$$

或

$$H_1: P(X > Y) > P(X < Y)$$

H_1 的 3 种形式都以不同的方式传达这样的思想，"X 趋向于比 Y 大".

如果 T 大于它的 $1-\alpha$ 分位数，则以水平 α 拒绝 H_0，分位数可从表 A7 和（3）式得到. 我们很容易发现 $T' = n(N+1) - T$，且若 T' 小于表 A7 中所给的 α 分位数，则拒绝 H_0. 如果 $n > 20$ 或 $m > 20$，则用（5）式来找分位数.

如果用 T_1，则若 $T_1 > z_{1-\alpha}$ 就拒绝 H_0，$z_{1-\alpha}$ 从表 A1 得到.

p-值近似于下面的概率

$$p\text{-值} \cong P\left(Z \geqslant \frac{T - \frac{1}{2} - n\frac{N+1}{2}}{\sqrt{\frac{nm(N+1)}{12}}}\right) \tag{8}$$

它可从表 A1 得到，并且与下面的公式相同

$$p\text{-值} \cong P\left(Z \leqslant \frac{T' + \frac{1}{2} - n\frac{N+1}{2}}{\sqrt{\frac{nm(N+1)}{12}}}\right) \tag{9}$$

如果利用 T_1，则 p-值可简单地写成：

$$p\text{-值} = P(Z \geqslant T_1) = 1 - P(Z \leqslant T_1) \tag{10}$$

275 它可直接从表 A1 获得.

计算机辅助 含有 Mann-Whitney 检验的计算机程序在 *Minitab*, *S-plus*, *SAS* 和 *StatXact* 中都可找到. ◄

评注 当检验上述包含 $P(X \geqslant Y)$ 的假设时，Mann-Whitney 检验是无偏的和相合的. 然而，对于包含 $E(X)$ 和 $E(Y)$ 的假设却不总是这样的. 为保证对含有 $E(X)$ 的假设其检验仍然相合且无偏，只要对前面的模型加入另一个假定即可.

假定 4. 如果总体分布函数之间有差异，而这种差异只是分布位置的差异. 即如果 $F(x)$ 和 $G(x)$ 不等，则 $F(x)$ 和 $G(x+c)$ 相等，其中 c 是某常数

例 5.1.1

某高中四年级有 48 个男生. 12 个男生住在农场，其余 36 个住在城镇. 设计一个检验，看是否来自农场的男生比来自城镇的男生健壮. 对这个班的每个男生做身体健壮测试，低分说明身体条件差. 农场男生(X_i)和城镇男生(Y_j)的得分如下.

X_i: 农场男生		Y_j: 城镇男生					
14.8	10.6	12.7	16.9	7.6	2.4	6.2	9.9
7.3	12.5	14.2	7.9	11.3	6.4	6.1	10.6
5.6	12.9	12.6	16.0	8.3	9.1	15.3	14.8
6.3	16.1	2.1	10.6	6.7	6.7	10.6	5.0
9.0	11.4	17.7	5.6	3.6	18.6	1.8	2.6
4.2	2.7	11.8	5.6	1.0	3.2	5.9	4.0

每组男生都不是来自任何总体的随机样本. 然而，合理的假定这些得分是来自此年龄组农场和城镇男生总体的组合随机样本，至少对类似的局部而言是这样. 模型其他的假设似乎都是合理的，例如两组之间的独立性. 因此，选择 Mann-Whitney 检验来检验

H_0: 农场男生不比城镇男生更健壮

H_1: 农场男生比城镇男生更健壮

这些假设建议用假设检验 C 所述的右边检验. 得分的秩赋予如下. 276

X	Y	秩	X	Y	秩	X	Y	秩
	1.0	1		6.2	17		11.3	33
	1.8	2	6.3		18	11.4		34
	2.1	3		6.4	19		11.8	35
	2.4	4		6.7	20.5	12.5		36
	2.6	5		6.7	20.5		12.6	37
2.7		6	7.3		22		12.7	38
	3.2	7		7.6	23	12.9		39
	3.6	8		7.9	24		14.2	40
	4.0	9		8.3	25		14.8	41.5
4.2		10	9.0		26	14.8		41.5
	5.0	11		9.1	27		15.3	43
	5.6	13		9.9	28		16.0	44
	5.6	13		10.6	30.5	16.1		45
5.6		13		10.6	30.5		16.9	46
	5.9	15	10.6		30.5		17.7	47
	6.1	16		10.6	30.5		18.6	48

有 4 组得分存在结，由方括号标明. 正如注明的那样，在每个组内将应赋的秩平均，并把平均秩赋给组内的每个值.

检验是单边的，临界域对应着大的 T_1 值. 注意到并没有很多结，所以用 T 替代 T_1 是可以接受的，在这个例子的后面将比较这两种方法. 从表 A1 可以得到，$\alpha=0.05$ 的临界域对应的 T_1 值大于 1.6449.

这里，我们有 $n=12, m=36$，所以，$N=12+36=48$. 赋给 X 的秩和是：

$$T=\sum_{i=1}^{n}R(X_i)$$

$$=6+10+13+18+22+26+30.5+34+36+39+41.5+45=321$$

这48个秩的平方和是：

$$\sum_{i=1}^{N}R_i^2=38{,}016$$

比从1到48（没有结）秩的平方和38,024稍微小一点（用引理1.4.2.）. 现在我们

277 计算 T_1.

$$T_1=\frac{T-n\dfrac{N+1}{2}}{\sqrt{\dfrac{nm}{N(N-1)}\sum_{i=1}^{N}R_i^2-\dfrac{nm(N+1)^2}{4(N-1)}}}$$

$$=\frac{321-12\dfrac{49}{2}}{\sqrt{\dfrac{(12)(36)}{(48)(47)}(38{,}016)-\dfrac{(12)(36)(49)^2}{4(47)}}}$$

$$=0.6431$$

不在临界域内，所以接受 H_0，我们得出结论：这些数据并不表明农场男生比城镇男生更健壮. 和 $T_1=0.6431$（由表A1得到）比较，发现0.6431接近于0.74分位数，所以以水平 α 为 $1-0.74=0.26$ 拒绝零假设，并且 p-值为0.26.

如果我们忽略这几个结，并用（5）式的大样本逼近，就可以得到 T 的近似0.95分位数为：

$$w_{0.95}=n\frac{N+1}{2}+(1.6449)\sqrt{nm(N+1)/12}$$

$$=294+(1.6449)(42)$$

$$=363.1$$

这和前面一样接受 H_0. ■

下一个例子将解释没有明确定义随机变量的情形. 根据直接比较相互之间的硬度排列燧石，给每块燧石赋一个硬度度量的随机变量是可以接受的，但在这种情形下并不必要

例5.1.2

设计了一个简单的试验，看A区的燧石硬度是否与B区的燧石硬度相同. 在A区收集了4块燧石，在B区收集了5块燧石. 为测定两块燧石中哪块更硬，用两块燧石互相摩擦，破损少的那块被判定是两块中更硬的. 用这种方式将9块燧石按硬度排

278 序，秩1赋给最软的一块，秩2赋给次最软的，等等.

原始燧石块	A	A	A	B	A	B	B	B	B
秩	1	2	3	4	5	6	7	8	9

要检验的假设是：

$$H_0\text{: A区和B区的燧石硬度相同}$$

备择假设是：

$$H_1\text{:燧石硬度不相同}$$

用 Mann-Whitney 双边检验，其中，$n=4$，$m=5$，且

$$\begin{aligned} T &= \text{A区石块的秩和} \\ &= 1+2+3+5 \\ &= 11 \end{aligned}$$

近似 0.05 的双边临界域对应着 T 值小于 12 和 T 值大于 $(4)(10)-12=28$. 由于在这个例子中 T 小于 12，零假设被拒绝，得出结论是两个区的燧石硬度不相同. 由于趋势的不同，进一步的结论是 B 区的燧石比 A 区的燧石更硬. p-值可以近似地从 (6) 式和表 A1 得到.

$$\begin{aligned} p\text{-值} &\cong 2\cdot P\left(Z\leqslant\frac{11+\frac{1}{2}-4\frac{10}{2}}{\sqrt{\frac{4\cdot 5\cdot 10}{12}}}\right) \\ &= 2\cdot P(Z\leqslant -2.0821) \\ &= 2(0.019)=0.038 \end{aligned}$$

■

□**理论**　假设 X_i 和 Y_j 是同分布的，可以找到 T 的零分布. 如果 X_i 和 Y_j 独立同分布，则在组合的样本中，X 和 Y 的每个排列都是等可能的，这是很多秩检验所依赖的基本原则. 对这个陈述的证明需要微积分学，因此超出了本书的范围. 然而，当人们试图提供一个理由说明一些排列比其他的更可能后，这一陈述的真实性看起来显得比较直观，对此没有更有效的理由了. 因此，作为直观明显而未证明的陈述，我们这里可以接受这个事实，即所有有序的排列是等可能的. 279

如果 X_i 和 Y；是独立同分布的，则在这个组合的样本中赋给 X_i 的秩应该与从 1 到 $n+m$ 整数中随机选择 n 个整数类似. 即没有理由说明为什么赋给 X_i 一个值某个指定秩的机会比赋给它其他任何秩的机会大. 由于 1 到 $n+m$ 中的每个数是等可能地赋给 X_i 作秩的，由于 n 个不同的数选为 X 的秩，秩和 T 的概率分布可以通过考虑无放回地从 1 到 $n+m$ 个整数中随机选取 n 个整数和的概率分布得到.

从 $n+m$ 个整数中选取 n 个整数的方式总共有 $\binom{n+m}{n}$ 种，根据刚刚阐述的基本前提，每种方式出现的概率相同. 因此，通过计算从 1 到 $n+m$ 中取不同的 n 个整数且和为 k 的个数，再用这个数除以 $\binom{n+m}{n}$，即为 $T=k$ 的概率.

例如，在例 5.1.2 中，如果样本大小 $n=4,m=5$，从 9 个秩中选出 4 个，不同方式的个数为

$$\binom{n+m}{n}=\frac{9!}{4!5!}=126$$

当1,2,3,4这4个秩被选到时，T的最小值可能是10. 仅当1,2,3,5被选到时，T的下一个值是11. $T=12$有两种方式，即秩为1,2,3,6时或秩为1,2,4,5时. 因此

$$\begin{aligned}
P(T=10)&=1/126 \qquad P(T\leqslant 10)=0.0079\\
P(T=11)&=1/126 \qquad P(T\leqslant 11)=0.0159\\
P(T=12)&=2/126 \qquad P(T\leqslant 12)=0.0317\\
&\text{等.} \qquad\qquad\qquad \text{等.}
\end{aligned}$$

注意，在例5.1.2中精确的p-值是2. $P(T\leqslant 11)$，这已经由上面给出，是$4/126=$
280 0.0317，与这个例子中用的近似p-值0.038差距不大.

由于T是n个X的秩和，对于大的n和m，可以用中心极限定理来得到T的渐近分布，这已在例1.5.7中完成，其结果表明T是渐近正态的，它的均值和方差(由定理1.4.5给出)：

$$E(T)=\frac{n(n+m+1)}{2} \tag{11}$$

和

$$\mathrm{Var}(T)=\frac{n(n+m+1)m}{12} \tag{12}$$

因此，根据定理1.5.1，T的分位数近似为

$$w_p=E(T)+z_p\sqrt{\mathrm{Var}(T)} \tag{13}$$

其中，z_p是标准正态分布的p分位数，除了$\mathrm{Var}(T)$必须基于两个样本中所用的真实秩和平均秩外，用正态逼近T_1的理由和前面用正态逼近T的理由相似，细节的讨论放在5.3节. □

Mann-Whitney 检验可以用于检验

$$H_0: E(X)=E(Y)+d, \qquad \text{或} \qquad E(X)-E(Y)=d \tag{14}$$

这里d是某个指定的数. 我们简单地将d加到每个Y_i上，然后将Mann-Whitney检验用到原始X和新调整的Y上.

导致接受上述H_0的所有d值的集合，就是我们所得到的两个期望差异$E(X)-E(Y)$的置信区间. 对试验者来说，这个置信区间比仅检验两个期望是否相等更有意义. 我们现在来介绍获得置信区间的一种方法，而不必一再运用Mann-Whitney检验.

▶两个期望差异的置信区间

数据 数据分别由样本大小为n和m的两个随机样本$X_1,\cdots,X_n$和$Y_1,\cdots,Y_m$组成.
281 令X和Y分别为与X_i和Y_j同分布的随机变量

假定条件

1. 两个样本都是来自各自总体的随机样本.
2. 除了每个样本内观测相互独立外，两个样本之间也相互独立.
3. 除了位置参数可能不同外，两个总体的分布函数相同. 即有一个常数d，使

得 X 和 $Y+d$ 有相同的分布函数.

注意，这里不需要假设连续性. Noether(1967b)证明如果从连续总体抽样时，置信区间的置信系数是 $1-\alpha$，则对于一般的总体，同样的置信区间包括端点时，置信系数至少是 $1-\alpha$，而不包括端点时至多是 $1-\alpha$. 我们这里将包含端点.

方法 对固定的 n 和 m，用表 A7 来确定 $\alpha/2$ 分位数 $w_{\alpha/2}$；如果 n 和 m 很大时，则用 (5) 式来确定，这里 $(1-\alpha)$ 是理想的置信系数. 注意，在有很多结存在时，表 A7 和 (5) 式也适用. 然后用下式计算 k

$$k = w_{\alpha/2} - n(n+1)/2 \tag{15}$$

从所有可能的数对 (X_i, Y_j) 中找出 k 个最大的 $X_i - Y_j$ 和 k 个最小的 $X_i - Y_j$. 为了找出最大差异和最小差异，先将每个样本从小到大排序是很方便的，然后用 X 的数据作行，Y 的数据作列，形成差异 $X_i - Y_j$ 的矩阵. 第 k 个最大的差异是上限 U，第 k 个最小的差异是下限 L，即，对所有 mn 个可能差异的有序排列，分别从排列的左右两边向中间数，则相应的第 k 个差异就是 L 和 U. 然后得到置信区间

$$P[L \leqslant E(X) - E(Y) \leqslant U] \geqslant 1-\alpha \tag{16}$$

计算机辅助 两个期望（或中位数）之差异的精确非参数置信区间可以从 *Minitab* 和 *StatXact* 中得到. 这些也是众所周知的位移 Hodges-Lehmann 估计. ◀

例 5.1.3

蛋糕糊达到一定的稠度将要被搅拌. 5 批糊状物将用搅拌器 A 来搅拌，另外 5 批用搅拌器 B 来搅拌. 搅拌的时间由下表给出（分钟）. 282

搅拌器A	7.3	6.9	7.2	7.8	7.2
搅拌器B	7.4	6.8	6.9	6.7	7.1

要求搅拌时间均值之差的 95% 置信区间，具体地说是 $E(X) - E(Y)$ 的 95% 置信区间，这里 X 是指搅拌器 A 的时间，Y 是指搅拌器 B 的时间.

对于 $n=5, m=5, \alpha=0.05$，从表 A7 得到 $w_{0.025}=18$，所以 $k=18-(5)(6)/2=3$. 把两个样本从小到大排序，X 作行，Y 作列，形成差异 $X_i - Y_j$ 的矩阵如下.

X_i \ Y_j	6.7	6.8	6.9	7.1	7.4
6.9	0.2	0.1	0.0	−0.2	−0.5
7.2	0.5	0.4	0.3	0.1	−0.2
7.2	0.5	0.4	0.3	0.1	−0.2
7.3	0.6	0.5	0.4	0.2	−0.1
7.8	1.1	1.0	0.9	0.7	0.4

于是，得到最大和最小差异.

最小差异	**最大差异**
$6.9 - 7.4 = -0.5$	$7.8 - 6.7 = 1.1$
$6.9 - 7.1 = -0.2$	$7.8 - 6.8 = 1.0$
$7.2 - 7.4 = -0.2 = L$	$7.8 - 6.9 = 0.9 = U$

所以得到 $E(X) - E(Y)$ 的 95% 置信区间为 (L, U)，即 $(-0.2, 0.9)$. ■

□**理论** 注意,有 mn 个对(X_i,Y_j). 记 k 为 $X_i>Y_j$,即 $X_i-Y_j>0$ 数据对的个数,则(1)式中的 T(X 的秩和)是 $k+n(n+1)/2$(见思考题1). 因为如果没有一个 Y 比 X 小,则 $T=1+2+\cdots+n=n(n+1)/2$(根据引理1.4.1),而有 k 对(X_i,Y_j)使得 Y 比 X 小的影响是 T 增大了 k 个单位.

使得 H_0 刚好接受的 T 的“边界”值在表A7中以 $w_{\alpha/2}$ 给出. 从 $w_{\alpha/2}$ 中减去 $n(n+1)/2$,我们就得到 k 的边界值. 现在我们要求 d 的值使得它加给 Y 后刚好能得到 k 的边界值,即使得恰好有 k 个数对(X_i,Y_j+d),满足 $X_i>Y_j+d$ 或 $X_i-Y_j>d$.

如果我们将所有差异 X_i-Y_j 中的最大值加到每个 Y 上,显然没有一个 X 会比调
283 整的 Y 大,因为这些 Y 太大了. 将差异 X_i-Y_j 中的第 k 个大值加到每个 Y 上,我们将得到边界情况:满足 $X_i>Y_j+d$ 的数对不到 k 个,满足 $X_i\geqslant Y_j+d$ 的数对至少有 k 个. 我们由此得到 d 的最大值,它使得能接受 $H_0:E(X)=E(Y)+d$. 将上述过程从左端开始再做一遍,我们得到 d 的最小值,它使得能接受相同的假设. 这样 d 值的集合给出了我们所求的置信区间. □

与其他方法比较

与Mann-Whitney检验相比的自然方法,就是早些提到的两样本 t 检验. 这种 t 检验的形式将两样本的样本均值 $\overline{X}$ 和 $\overline{Y}$ 包含在下面的公式中.

$$t=\frac{(\overline{X}-\overline{Y})\sqrt{mn(N-2)/N}}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2+\sum_{j=1}^{m}(Y_j-\overline{Y})^2}} \tag{17}$$

t 值要与自由度为 $N-2$ 的 t 分布的分位数(由表A21获得)比较. 为了使这些分位数比较精确,必须另外假设两个总体都是正态分布. 有了这个假设,t 检验是最强功效的检验. 当 t 检验与Mann-Whitney检验相比时,一些非正态型分布会导致检验有很小的功效,特别是当一个或两个样本中都有异常的大或小的观测(称为“离群值”)时,是这样.

如果 t 统计量的计算机程序是可行的,则它可以用来简化Mann-Whitney检验中的计算,尤其是有很多结时. 因为仅要计算基于秩 $R(X_i)$ 和 $R(Y_j)$ 而不是 X_i 和 Y_j 的 t 统计量,用这个结果与自由度为 $N-2$ 的 t 分布的分位数(可由表A21获得)做比较. 尽管这个逼近和通常的正态逼近不尽相同,但它在大多数情况下更精确. 对于一个甚至更好的逼近来说,求由(2)式给出的 T_1 的平均值,计算基于秩的 t 统计量,并将它与从表A1和表A21得到的两个分位数的平均做比较. 这个方法更多的细节可参看Iman(1976).

在假设X和Y的分布除了它们的均值外是相同的前提下,我们考虑Mann-Whitney检验与t检验相比的渐近相对效率(A. R. E.).如果总体是正态的,则A. R. E.是0.955;如果总体是均匀的,则A. R. E. 是1.0;如果总体是对称的双指数分布,则A. R. E. 是1.5,如果两个总体只是位置参数不同,则A. R. E. 不会小于0.864,

但是可能会达到无穷大（Hodges 和 Lehmann，1956）. 284

中位数检验也可以用于这种类型的数据. Mann-Whitney 检验相对于中位数检验的 A. R. E. 而言，对于正态分布是 1.5，对于均匀分布是 3.0，但对于双指数分布只有 0.75. 记住这是渐近（asymptotic）相对效率. 对于小样本，Mann-Whitney 检验在双指数分布中可能比中位数检验中的功效更高（见 Conover, Wehmanen 和 Ramsey, 1978）. 另一方面，当 H_0 为真时，中位数检验并不要求分布相同，它只要求它们有相同的中位数. 因此中位数检验可以用到 Mann-Whitney 检验无效的情形中.

Mann-Whitney 检验首先是由 Wilcoxon(1945) 在 $n = m$ 的情形下介绍的. Wilcoxon 检验由 White(1952) 和 van der Reyden(1952) 推广到样本大小不等的情形. Festinger(1946) 独立发展和介绍了与 Wilcoxon 的检验等价的形式. Mann 和 Whitney 好像是最先考虑了样本大小不等的情形，并提供了适合小样本使用的表. 很大程度上是因为 Mann 和 Whitney 的工作引发了这个检验的广泛使用. 由于这个检验归属于不同的作者，叫它哪个名字是使用者的特权.

为了检查散布或方差或尺度的不同，Siegel 和 Tukey 在 1960 年介绍了对 Mann-Whitney 检验所做的修正，它和 Freund 和 Ansari(1957) 设计的早些的检验在原则上是相似的. 这两个检验之间的关系在 Hájek 和 Sidák(1967) 的第 126 页有描述.

对于 Mann-Whitney 检验更广泛的表，Verdooren(1963) 给出了 n 和 $m \leqslant 25$ 情形下的表，Milton(1964) 给出了 $n \leqslant 20$ 和 $m \leqslant 40$ 情形下的表. 其他表格和参考书可以在 Jacobson(1963) 中找到. Klotz(1966) 和 Buckle, Kraft 和 van Eeden(1969) 讨论过 Mann-Whitney 检验统计量的分布. 其他参考文章可见 Zaremba(1965) 和 Serfling(1968).

Mann-Whitney 和其他密切相关检验的效率是 Chanda(1963), Noether(1963), Hayman 和 Govindarajulu(1966), McNeil(1967), R. A. Shorack(1967), Stone(1967) 以及 Conover 和 Kemp(1976) 的文章的主题. Conover(1973a) 给出了结的处理理由. Alling(1963), Woinsky 和 Kurz(1969), Bradley, Martin 和 Wilcoxon(1965), Bradley, Merchant, Wilcoxon (1966), Sen 和 Ghosh(1974) 以及 Spurrier 和 Hewitt(1976) 给出了后续检验的修正. Batschelet(1965) 讨论了检验圆分布的问题，并由 Beran(1969) 和 Schach(1969b) 解决.

如果两个样本是删失的（即如果一些最大的/或最小的样本值是不可观测的），则数据可以用修正的 Mann-Whitney 检验来分析，正如 Gastwirth(1965a), Gehan(1965a, 1965b), Gehan 和 Thomas(1969), Saw(1966), Basu(1968), Hettman-sperger(1968) 和 Shorack (1968) 讨论的那样. Mardia(1967a, 1968) 给出了二维两样本秩检验问题. 其他两样本非参数检验由 Hudimoto(1959), Haga(1960), Tamura(1963), Potthoff (1963), Wheeler 和 Watson (1964), Gastwirth(1965b), Bhattacharyya 和 Johnson (1968), Mielke (1972) 和 Pettitt (1976) 提出并讨论过. Mikulski(1963), Basu 285
(1967a), Hollander (1967a) 以及 Gibbons 和 Gastwirth(1970) 验证了一些检验的效率. 其他相关的文章包括 Hollander, Pledger 和 Lin(1974), Bickel 和 Lehmann(1975), Hettmansperger 和 Malin(1975), Doksum 和 Sievers(1976) 以及 Fligner, Hogg 和 Killeen(1976).

Noether(1967a)讨论过寻找置信区间的方法，Walker 的 Moses 和 Lev(1953)描述过相关的图方法. 当样本容量较大时，McKean 和 Ryan(1977)给出了可能有用的算法. 其他位置差异的估计由 Hodges 和 Lehmann(1963),Høyland(1965),Rao,Schuster 和 Littell(1975)以及 Switzer(1976)讨论过. 相关的文章还有 Moses(1965),Govindarajulu(1968),Bauer(1972),Ury(1972)以及 Kraft 和 van Eeden(1972).

习题

1. 检验下面的数据，看 Des Moines 市的平均高温是否比 Spokane 市的平均高温高，数据是在夏天随机抽样得出的.

Des Moines	83	91	94	89	89	96	91	92	90
Spokane	78	82	81	77	79	81	80	81	

2. 在一个可控环境实验室中，检验 10 位男士和 10 位女士，来决定他们认为最适合的室内温度. 结果如下

男士	74	72	77	76	76	73	75	73	74	75
女士	75	77	78	79	77	73	78	79	78	80

假设这些温度好似来自各自总体的随机样本，问男士和女士平均适合的温度相同吗?

286 3. 用现有的方法教 7 个学生学代数，而 6 个学生用新方法来学习代数. 求用这两种方法学习所得成绩分数差异的 90% 置信区间.

方法	**学生成绩分数**						
现有的	68	72	79	69	84	80	78
新的	64	60	68	73	72	70	

4. 把食物 A 给了 4 个超重的女孩，把食物 B 给了其余 5 个超重的女孩，所观测到的减肥重量如下. 求出两种食物平均效率差异的 90% 置信区间.

食物	**减肥重量(磅)**
A	7, 2, −1, 4
B	6, 5, 2, 8, 3

5. 一个试验中的 8 名志愿者随机分为两组，看望远镜瞄准器是否能提高微光条件下射中目标的能力. A 组给了带有望远镜瞄准器的步枪，B 组有同样的步枪，但带的是开放式瞄准器. 学习了一段时间后，对他们进行了微光下射击测验. 下面是他们的分数（满分 100 分）

A组	96	93	88	85
B组	89	83	80	77

你能得到什么结论?

6. 在一个树木繁茂的地方建立起 10 个普通伪装的帐篷和 10 个有图案伪装的帐篷，一队观测者出发去寻找它们，并报告他们第一眼看到每个帐篷的距离（仅真正看到的），直到所有 20 个帐篷都被找到为止. 这个研究的目的是确定有图案伪装是否比普通伪装更难发现. 每个帐篷被发现的距离如下表

伪装类型	**距离(米)**
普通的	25, 28, 16, 34, 38, 21, 29, 43, 32, 36
带图案的	26, 12, 16, 21, 20, 14, 10, 18, 22, 20

(a) 作一个假设检验.

(b) 求出平均发现距离差异的 95% 非参数置信区间.

思考题

1. 设 S 为 X_i大于 Y_j 的 (X_i, Y_j) 的对数（结计数为0.5）. 注意到总共有 mn 对. 证明 S 和 T 满足关系

$$S = T - \frac{n(n+1)}{2}$$

作为 $P(X > Y)$的估计，哪个统计量看起来是合理的? 287

2. 在 $n=3, m=2$，H_0 为真的情况下，求出 T 的精确分布，并和表 A7 比较.
3. 计算习题 2 中数据的两样本 t 统计量（（17）式），并把结果和用 Mann-Whitney 检验得到的结果进行比较.

5.2 多个独立样本

Kruskal 和 Wallis(1952)把 5.1 节中提出的两个独立样本的 Mann-Whitney 检验推广到分析 $k(k>2)$个独立样本的问题. 试验情形是 k 个随机样本已经得到，且来自 k 个可能不同的总体，我们要检验零假设：所有总体分布都相同，对备择假设：有些总体提供比其他总体偏大的观测值. “偏大”是运用到随机变量的观测值上，但实际上，这些观测值可以是根据一些诸如质量，值之类的性质按升序排列的任何观测值，它们可以用 Kruskal-Wallis 检验来分析，这个分析过程类似于 Mann-Whitney 检验分析非数值数据的方式，正如例 5.1.2 中的那样.

▶ Kruskal-Wallis 检验

数据 数据由 k 个样本容量可能不同的随机样本组成. 记第 i 个样本容量为 n_i 的随机样本为 $X_{i1}, X_{i2}, \cdots, X_{in_i}$，则数据可以排成许多列.

样本 1	样本 2	· · ·	样本 k
$X_{1,1}$	$X_{2,1}$		$X_{k,1}$
$X_{1,2}$	$X_{2,2}$		$X_{k,2}$
· · ·	· · ·		· · ·
X_{1,n_1}	X_{2,n_2}		X_{k,n_k}

记 N 为观测的总数

$$N = \sum_{i=1}^{k} n_i \tag{1}$$

给 N 个观测值中最小的赋秩 1，给第二小的赋秩 2，等等，N 个观测中最大的赋秩 N. 令 $R(X_{ij})$代表赋给 X_{ij}的秩，令 R_i 为赋给第 i 个样本的秩和，即 288

$$R_i = \sum_{j=1}^{n_i} R(X_{ij}) \qquad i = 1, 2, \ldots, k \tag{2}$$

对每个样本计算 R_i.

如果有几个观测互相相等，则可能会有不同的方式赋秩，像本章前面的检验中

那样，给每个有结的观测赋平均秩.

假定条件

1. 所有样本均为来自各自总体的随机样本.
2. 除了每个样本内观测相互之间独立外，每个样本之间也是相互独立的.
3. 度量尺度至少是须序的.
4. 或者 k 个总体分布函数相同，或者一些总体产生比另一些总体产生大的值.

检验统计量　检验统计量定义为

$$T=\frac{1}{S^2}\left(\sum_{i=1}^{k}\frac{R_i^2}{n_i}-\frac{N(N+1)^2}{4}\right) \tag{3}$$

其中，N 和 R_i 分别在（1）式和（2）式中定义，

$$S^2=\frac{1}{N-1}\left(\sum_{\text{所有秩}}R(X_{ij})^2-N\frac{(N+1)^2}{4}\right) \tag{4}$$

如果没有结，则 S^2 简化为 $N(N+1)/12$，检验统计量简化为

$$T=\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}-3(N+1) \tag{5}$$

如果结的个数是中等的，（3）式和（5）式之间差别很小，所以我们更愿意使用简化的（5）式.

零分布　对于 $k=3$ 和所有的 $n_i\leqslant 5$，表 A8 给出了 T 的精确分布，但是一般情形下，精确
289 分布相当难求. 因此，自由度为 $k-1$ 的 χ^2 分布将作为 T 的零分布逼近.

假设

H_0: 所有的 k 个总体分布函数相同

H_1: 至少有一个总体会产生比其他至少一个总体偏大的观测

因为 Kruskal-Wallis 检验的设计对检验 k 个总体之间均值差异是敏感的，所以备择假设有时也有如下叙述.

H_1: k 个总体没有相同的均值

如果 T 大于它零分布的 $1-\alpha$ 分位数，则以水平 α 拒绝 H_0. 如果 $k=3$，所有的样本容量为 5 或更小，而且没有结，则精确的分位数可以从表 A8 中得到. Iman，Quade 和 Alexander(1975)中给出了更广泛而精确的表. 当有结存在时，或者当精确表不可用时，近似的分位数（即自由度为 $k-1$ 的 χ^2 分布的分位数）可以从表 A2 中得到. 如果 T 大于这样得到的 $1-\alpha$ 分位数，则以水平 α 拒绝 H_0. p-值近似为自由度为 $k-1$ 的 χ^2 分布随机变量超过 T 观测值的概率.

多重比较　当且仅当零假设被拒绝时，我们用下面的方法决定总体的哪些对不同. 如果下面的不等式满足，我们可以说总体 i 和 j 不同：

$$\left|\frac{R_i}{n_i}-\frac{R_j}{n_j}\right|>t_{1-(\alpha/2)}\left(S^2\frac{N-1-T}{N-k}\right)^{1/2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)^{1/2} \tag{6}$$

其中，R_i 和 R_j 是两个样本的秩和，$t_{1-\alpha/2}$ 是从表 A21 得到的自由度为 $N-k$ 的 t 分布的（$1-\alpha/2$）分位数，S^2 由（4）式确定，T 由（3）式或（5）式确定. 可对总体

的所有对重复使用这个方法. 正如在 Kruskal-Wallis 检验中所用到的 α 一样，这里所用的是同样的 α 水平.

计算机辅助 含有 Kruskal-Wallis 检验的计算机程序在 *Minitab*, *S-Plus*, *SAS* 和 *StatXact* 中都有. 这些和其他程序都会将数据转换为秩，并作秩的单因素方差分析，正如思考题 5 中所讨论的. 这个方法自动修正结，而且包含更宽可供选择的多重比较方法. ◀

例 5.2.1

数据来自例 4.3.1 给出的完全随机化的设计，其中种玉米的 4 种不同方法导致了不同地块上每亩产量的不同，通常只用一个统计分析，但是这里我们用 Kruskal-Wallis 检验，使得一个困难的比较可以用中位数检验来处理，它先提供一个比 0.001 稍小一点的 p-值. 290

假设表述如下：

H_0: 4种方法等价

H_1: 一些种玉米的方法比其他方法的产量更高

把观测从最小的 77，秩为 1，到最大的 101，秩为 $N=34$ 排列，有结的值赋平均秩. 观测的秩以及秩和 R_i 给出如下.

方法

	1		2		3		4	
	观测	秩	观测	秩	观测	秩	观测	秩
	83	11	91	23	101	34	78	2
	91	23	90	19.5	100	33	82	9
	94	28.5	81	6.5	91	23	81	6.5
	89	17	83	11	93	27	77	1
	89	17	84	13.5	96	31.5	79	3
	96	31.5	83	11	95	30	81	6.5
	91	23	88	15	94	28.5	80	4
	92	26	91	23			81	6.5
	90	19.5	89	17				
			84	13.5				
R_i:		196.5		153.0		207.0		38.5
n_i:		9		10		7		8
$N=34$								

近似大小 $\alpha=0.05$ 的临界域对应着大于自由度为 $k-1=3$ 的 χ^2 分布随机变量的 0.95 分位数的 T 值，由表 A2 可得，它为 7.815（注意，中位数检验也用自由度为 $k-1$ 的 χ^2 分布，所以尽管检验统计量不同，这两个检验的临界域将会相同.）

用（5）式得到的 T 值为：

$$T = 25.46$$

很显然导致拒绝 H_0. 与中位数检验比较，Kruskal-Wallis 检验功效的粗略想法可由比较两个检验统计量的值得到. 两个检验统计量有相同的渐近分布，即自由度为 3 的 χ^2 分布. 但是由 Kruskal-Wallis 检验得到的值 25.46 比由中位数检验计算的值 17.6 大得多，表明了它对样本差异更敏感. 291

因为拒绝了 H_0，所以可用多重比较的方法．我们忽略不多的结，并用下面的简单形式

$$S^2 = N(N+1)/12 = 99.167 \tag{7}$$

使得

$$\frac{S^2(N-1-T)}{N-k} = \frac{(99.167)(33-25.464)}{34-4} = 24.911 \tag{8}$$

剩下的计算如下

总体	$\left\|\frac{R_i}{n_i} - \frac{R_j}{n_j}\right\|$	$2.041(24.911)^{1/2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)^{1/2}$
1 和 2	6.533	4.681
1 和 3	7.738	5.134
1 和 4	17.021	4.950
2 和 3	14.271	5.020
2 和 4	10.488	4.832
3 和 4	24.759	5.272

在每个情况中，第二列的数都超过了第三列，所以我们可以说，多重比较方法表明总体中的每对都不同． ■

对有很多结的情形，应当毫不犹豫地应用本章中的秩检验．事实上，Kruskal-Wallis 检验是用于列联表的一个非常好的检验，如下表所示，行代表有序分类，列代表不同的总体．

	总体 1	2	3	···	k	行总和	$\overline{R}_i$ = 平均秩
类 1	O_{11}	O_{12}	O_{13}	···	O_{1k}	t_1	$(t_1+1)/2$
2	O_{21}	O_{22}	O_{23}	···	O_{2k}	t_2	$t_1 + (t_2+1)/2$
3	O_{31}	O_{32}	O_{33}	···	O_{3k}	t_3	$t_1 + t_2 + (t_3+1)/2$
···	···	···	···	···	···		
c	O_{c1}	O_{c2}	O_{c3}	···	O_{ck}	t_c	$\sum_{i=1}^{c-1} t_i + (t_c+1)/2$
292 列总和	n_1	n_2	n_3	···	n_k	N = 全体总和	

Q_{ij}是总体 j 中落入第 i 类的观测个数．i 行的平均秩是 $\overline{R}_i$，如上表所示，它是从行总和计算出的．这个结构和普通列联表之间的差别是，类（行）是有序的，即第 1 行的所有观测值认为是互相相等的，但比第 2 行的观测值小，等等．为了计算检验统计量，我们推荐用以下形式．记 R_j 为总体（对应列）j 的秩和，

$$R_j = \sum_{i=1}^{c} O_{ij}\overline{R}_i \tag{9}$$

并用下面的等式计算 S^2，

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{c} t_i \overline{R}_i^2 - N(N+1)^2/4\right] \tag{10}$$

则如同前边，将（9）式和（10）式代到前面的（3）式中计算检验统计量 T. 注意到（10）式和（4）式产生同样的 S^2 值，但是（10）式在这个情形下更容易使用. 如果拒绝零假设，正如前面所描述的那样，就可以用多重比较方法来查明差异在哪里.

例 5.2.2

比较上学期 3 名教师给出的成绩，看是否有些教师给的成绩比其他教师给的偏低. 零假设是：

H_0: 3个老师所给的成绩相互一致

感兴趣的备择假设是：

H_1: 有些教师给的成绩比其他教师给的偏低

所要检验的成绩分数如下，

	教师				
成绩	1	2	3	行总和	平均秩
A	4	10	6	20	10.5
B	14	6	7	27	34
C	17	9	8	34	64.5
D	6	7	6	19	91
F	2	6	1	9	105
学生总数	43	38	28	109	

从（9）式可求得列秩和，

$R_1 = 2370.5 \qquad R_2 = 2156.5 \qquad R_3 = 1468$ 293

检查我们到目前的计算，的确如此，对 $N = 109, R_j$ 的和应该等于 $N(N+1)/2 = 5995$. 从（10）式我们计算得到 $S^2 = 941.71$，最后由（3）式可得 $T = 0.3209$.

水平为 0.05 的临界域对应着 T 值大于 5.991（自由度为 2 的 χ^2 分布的 0.05 分位数，由表 A2 获得），很明显接受零假设. 在所列数据基础上，没有一位老师打分比其他老师偏高或者偏低. ■

□ **理论** 假设所有的观测是来自于相同或同分布的总体，那么我们可得到 T 的精确分布. 方法是随机化，这个方法也用于求 Mann-Whitney 检验统计量的分布. 即在前面的假设下，合并大小分别为 $n_1, n_2, \cdots, n_k$ 的组，则秩从 1 到 N 的每一个排列都是等可能的，且以概率 $n_1!\,n_2!\cdots n_k!/N!$ 发生，它是将 N 个秩分为大小为 $n_1, n_2, \cdots, n_k$ 组的所有方式个数的倒数. 对每一个排列，计算 T 值. 把所有对应 T 值相等的概率相加，从而给出了 T 的分布.

例如，如果在三样本情形下，$n_1 = 2, n_2 = 1, n_3 = 1$，4 个秩有 12 种等可能的排列；因此每个排列的概率为 1/12. 这 12 种排列与对应的 T 值如下.

	样本			
排列	**1**	**2**	**3**	T
1	1, 2	3	4	2.7
2	1, 2	4	3	2.7
3	1, 3	2	4	1.8
4	1, 3	4	2	1.8
5	1, 4	2	3	0.3
6	1, 4	3	2	0.3
7	2, 3	1	4	2.7
8	2, 3	4	1	2.7
9	2, 4	1	3	1.8
10	2, 4	3	1	1.8
11	3, 4	1	2	2.7
12	3, 4	2	1	2.7

因此，对 $n_1=2, n_2=1, n_3=1$，其概率函数 $f(x)$ 和分布函数 $F(x)$ 如下，

x	$f(x)=P(T=x)$	$F(x)=P(T\leqslant x)$
0.3	$2/12=1/6$	$1/6$
1.8	$4/12=1/3$	$1/2$
2.7	$6/12=1/2$	1.0

294

T 的分布的大样本逼近是基于以下事实：(2) 式中的 R_i 是 n_i 个随机变量的和，所以对于较大的 n_i，我们可以用中心极限定理，因此当 H_0 为真时，

$$\frac{R_i-E(R_i)}{\sqrt{\mathrm{Var}\,(R_i)}}$$

近似地服从标准正态随机变量的分布. 由定理 1.4.5，R_i 的均值和方差表达如下

$$E(R_i)=\frac{n_i(N+1)}{2} \tag{11}$$

和

$$\mathrm{Var}\,(R_i)=\frac{n_i(N+1)(N-n_i)}{12} \tag{12}$$

因此，

$$\left[\frac{R_i-E(R_i)}{\sqrt{\mathrm{Var}\,(R_i)}}\right]^2=\frac{\{R_i-[n_i(N+1)/2]\}^2}{n_i(N+1)(N-n_i)/12} \tag{13}$$

近似于自由度为 1 的 χ^2 随机变量的分布. 如果 R_i 相互独立，则和

$$T'=\sum_{i=1}^{k}\frac{\{R_i-[n_i(N+1)/2]\}^2}{n_i(N+1)(N-n_i)/12} \tag{14}$$

的分布可以用自由度为 k 的 χ^2 分布来逼近. 但是，R_i 的和是 $N(N+1)/2$，故 R_i 之间是不独立的. Kruskal(1952)证明了如果用 $(N-n_i)/N$ 乘以 T' 的第 i 项，$i=1,2,\cdots,k$，则结果

$$T=\sum_{i=1}^{k}\frac{\{R_i-[n_i(N+1)/2]\}^2}{n_i(N+1)N/12} \tag{15}$$

是渐近于自由度为 $k-1$ 的 χ^2 随机变量的分布.（15）式仅是（5）式中各项的重排，而（5）式是检验统计量 T 的原始定义. 因此，我们对 Kruskal-Wallis 检验统计量的分布合理地运用了 χ^2 逼近. □

Kruskal 和 Wallis(1952)发现，对于小的 α（大约小于 0.10）和选择小的 n_1,n_2,n_3 的值，真正的显著性水平小于所陈述的 χ^2 分布分位数所给的显性著水平，这表明 χ^2 分布逼近在很多（但不是绝大多数）情况下提供了一个保守的检验. Gabriel 和 Lachenbruch(1969)指出，尽管样本容量可能很小，但 χ^2 逼近还是好的. Iman 和 Davenport(1976)将 χ^2 逼近和其他逼近作了比较. 295

对于两样本情形，Kruskal-Wallis 检验和 Mann-Whitney 检验是等价的. 回顾一下（5.1 节），在 Mann-Whitney 检验中，一样本是 $X_1,X_2,\cdots,X_n$，而另一个样本是 $Y_1,Y_2,\cdots,Y_m$. 统计量 T 由（5.1.1）式定义如下：

$$T=\sum_{i=1}^{n}R(X_i) \tag{16}$$

即是 X 在联合样本中的秩和，它对应于 Kruskal-Wallis 检验中的 R_1. Mann-Whitney 双边检验就是，如果统计量 T 太大或太小，则拒绝 H_0. 由于当样本容量很大时，T 近似于正态，则根据定理 1.5.3，若

$$\frac{T-E(T)}{\sqrt{\mathrm{Var}(T)}} \tag{17}$$

在合适的标准正态分位数之上或之下，或者若它的平方，

$$\frac{[T-E(T)]^2}{\mathrm{Var}(T)} \tag{18}$$

在自由度为 1 的 χ^2 分布的 $1-\alpha$ 分位数之上，我们就可以拒绝 H_0. 所以，我们把（18）式中的量作为检验统计量，则自由度为 1 的 χ^2 分布就可以用于 Mann-Whitney 双边检验中. 两样本的 Kruskal-Wallis 检验也利用自由度为 1 的 χ^2 分布来检验与 Mann-Whitney 双边检验中相同的假设. 事实上，Kruskal-Wallis 检验统计量与（18）式所给出的 Mann-Whitney 检验统计量的形式相同. 这一点的证明留给读者作为习题.

Conover(1973a)给出了不连续分布中使用秩检验的论述. 当出现结时，Klotz 和 Teng(1977)讨论了检验统计量的精确分布. 多重比较方法是简单的普通参数方法，称为 Fisher 最小显著差异，Conover 和 Iman(1979)对此有过描述，它用秩来计算，而不是用数据. 296

普通参数方法称为“单因素方差分析，”或者有时简称为单因素 F 检验，所用的统计量如下：

$$F=\frac{\left(\sum_{i=1}^{k}T_i^2/n_i-C\right)/(k-1)}{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}X_{ij}^2-\sum_{i=1}^{k}T_i^2/n_i\right)/(N-k)} \tag{19}$$

其中，T_i 是第 i 个样本中观测的和，$C = T^2/N$，T 是所有观测的总数．如果 Kruskal-Wallis 检验的假设有效，并且如果总体以正态分布作为共同的分布，那么 F 统计量的分位数可由表 A22 给出．查看 $k_1 = k-1$ 的列与标记有 $k_2 = N-k$ 的行，N 和 k 由试验给出．当零假设为真时，违反正态假设通常会对 F 统计量有一些影响．然而，当 H_0 为假时，在某些非正态的分布类型中，F 检验的功效可能比 Kruskal-Wallis 检验小很多．例如，包含离群值的数据更适合用 Kruskal-Wallis 检验．

相对于 F 检验，Kruskal-Wallis 检验的渐近相对效率（A. R. E.）从来不会小于0.864，但如果分布函数有相同的形状，只是均值不同，则它可能会是无穷大．如果总体是正态的，则 A. R. E. 是 $3/\pi = 0.955$；对于均匀分布，则相对于 F 检验的 A. R. E. 是1.0；对双指数分布，它是1.5．与中位数检验相比较，对刚才提到的3种分布，Kruskal-Wallis 检验的 A. R. E. 分别是1.5,3.0和0.75.

类似于 Kruskal-Wallis 检验，Steel（1960），Sherman（1965）以及 McDonald 和 Thompson（1967）用秩和检验来作多重比较．Tobach, Smith, Rose 和 Richter（1967）提供了作多重比较的一些表．Rizvi 和 Sobel（1967），Sobel（1967），Rizvi, Sobel 和 Woodworth（1968）以及 Puri 和 Puri（1969）描述了挑选最好总体的方法．对删失数据的秩检验由 Basu（1976b）和 Breslow（1970）提出；检验有序备择假设的秩检验由 G. R. Shorack（1967），Odeh（1971，1972）以及 Tryon 和 Hettmansperger（1973）提出；协方差分析的秩检验由 Puri 和 Sen（1969a）提出．其他关于秩检验和几个独立样本的工作见 Sen（1962，1966），Matthes 和 Truax（1965），Quade（1966），Crouse（1966），Sen 和 Govindarajulu（1966），Odeh（1967），Deshpande（1970）以及 Bhapkar 和 Deshpande（1968）．Quade（1967）讨论了协方差分析．Brunden（1972）考虑用秩来分析 2×3 列联表．

习题

297 1. 检验来自3个不同类型的电灯泡的随机样本，看灯泡能亮多久，结果如下．

牌子

A	B	C
73	84	82
64	80	79
67	81	71
62	77	75
70		

这些结果表明这几个牌子之间有显著差别吗？如果有，哪些牌子不同？

2. 对20个新雇员试用4个工作培训项目，每个培训项目随机分配给5个雇员．20个雇员在相同的管理人员安排下进行．在某一指定的时期结束后，管理人员根据雇员的工作能力对他们进行排序，将最小的秩赋给工作能力最低的雇员．

项目	秩
1	4, 6, 7, 2, 10
2	1, 8, 12, 3, 11
3	20, 19, 16, 14, 5
4	18, 15, 17, 13, 9

这些数据表明不同的培训项目的效率有差异吗？如果有，它们可能是哪些？

3. 在很多不同的农场检查由水和风造成农场土地的破坏，同时也记录了在每个农场上实施耕作的类型，结果如下.

	耕作类型			
	最小限度耕作	等高地形	梯田	其他
破坏程度	农场数			
没有破坏	17	19	4	21
轻微破坏	3	10	4	42
中等破坏	0	2	2	34
严重破坏	0	0	2	6

耕作类型影响破坏程度吗？如果是，哪些耕作类型有显著差异？

4. 由同一公司生产的 3 种不同类型的收音机，都有一年的保质期. 下表记录了多少个收音机需要替换，多少个是可修理的，或者多少个在保质期内没有退回的数据. 298

	类型		
	A	B	C
替换的	12	3	6
修理的	10	8	7
未退回的	82	96	58

不同收音机类型的可信赖度看起来有显著差别吗？如果有，哪些看起来有差别？

5. 在指定的一段时间内，给白鼠喂 5 种食物中的一种后，测量白鼠肝脏内铁的吸收量，5 种食物中的每一种都随机地分配给了 10 只白鼠.

食物A	食物B	食物C	食物D	食物E
2.23	5.59	4.50	1.35	1.40
1.14	0.96	3.92	1.06	1.51
2.63	6.96	10.33	0.74	2.49
1.00	1.23	8.23	0.96	1.74
1.35	1.61	2.07	1.16	1.59
2.01	2.94	4.90	2.08	1.36
1.64	1.96	6.84	0.69	3.00
1.13	3.68	6.42	0.68	4.81
1.01	1.54	3.72	0.84	5.21
1.70	2.59	6.00	1.34	5.12

不同食物会影响白鼠肝脏内铁的吸收量吗？

6. 把3个减肥计划的每一个分配给了12名志愿者，志愿者被分配到哪个计划是随机的，总共有36位志愿者，假设他们是来自可能要试用一种减肥计划人群中的随机样本. 检验零假设：在3种计划下减肥量的概率分布没有差异，备择假设是：在3种计划下减肥量的概率分布有差异. 每个人减掉的磅数结果如下.

计划A		计划B		计划C	
2	17	17	5	29	5
12	4	15	6	3	25
5	25	3	19	25	32
4	6	19	4	28	24
26	21	5	9	11	36
8	6	14	7	7	20

思考题

1. 证明在没有结存在时，（3）式和（5）式是等价的.

299 2. 在 $n_1=3, n_2=2, n_3=1$，且没有结的情形下，求当 H_0 为真时，Kruskal-Wallis 检验统计量的精确分布. 将你的结果与表A8给出的分位数进行比较.

3. 在两样本情形中，我们更喜欢用 Mann-Whitney 检验而不是 Kruskal-Wallis 检验的原因有哪些?

4. 证明（10）式和（4）式是等价的.

5. 假定（19）式中的 F 统计量是用秩 $R(X_{ij})$，而不是用观测值 X_{ij} 计算的，那么证明以下由（3）式给出的关于 F 和 T 的关系式

$$F=\frac{T/(k-1)}{(N-1-T)/(N-k)}$$

成立. 因此，如果用秩计算 F，则当 T 值较大时拒绝 H_0 与当 F 值较大时拒绝 H_0 是等价的.

5.3 等方差检验

几个总体比较的通常标准是基于均值或总体其他位置的度量. 然而在某些情形下，总体的方差或许是感兴趣的量. 例如，已有断言称：碘化银撒在云里的作用是提高导致降雨的方差. 这样的断言可以用本节提出的方法来检验.

方差的检验类似于上节提出的均值的检验，即检验 $H_0:E(X)=E(Y)$，将两个独立的样本合并、排序，并用 X 的秩和作为检验统计量. 回想一下 X 的方差的定义，它为 $(X-\mu)^2$ 的期望值，此处 μ 是 X 的均值. 因此要检验 $E[(X_i-\mu_x)^2]=E[(Y_j-\mu_y)^2]$，一个合理的做法是，记录来自两个独立样本的 $(X_i-\mu_x)^2$ 和 $(Y_j-\mu_y)^2$ 的值，并给它们赋予秩，再用 $(X_i-\mu_x)^2$ 的秩和作为检验统计量. Talwar 和 Gentle (1977) 研究了这个方法. 尽管这个方法可以用，但给秩先平方后相加会得到更大的功效. 本节将对这样一个检验进行更确切的描述.

▶方差的平方秩检验

数据 数据是由两个随机样本组成的，记 $X_1,X_2,\cdots,X_n$ 为来自总体 1，容量为 n 的随机样本，$Y_1,Y_2,\cdots,Y_m$ 为来自总体 2、容量为 m 的随机样本. 将每个 X_i 和 Y_j 转换为它到均值的绝对离差

$$U_i=|X_i-\mu_1|, i=1,...,n \tag{1}$$ 300

和

$$V_j=|Y_j-\mu_2|, j=1,...,m \tag{2}$$

其中，μ_1 和 μ_2 是总体 1 和 2 的均值. 如果 μ_1 和 μ_2 未知，用 $\bar{X}$ 代替 μ_1，$\bar{Y}$ 代替 μ_2，则以下的检验仍然近似有效.

以通常方式将秩 1 到 $n+m$ 赋给 U 和 V 的合并样本. 如果 U 和/或 V 的几个值确实互相相等（存在结），则给它们的每个值都赋以没有结时要赋给它们的秩的平均值，记 $R(U_i)$ 和 $R(V_j)$ 为赋以的秩或平均秩. 注意，对 U_i 和 V_j 的排序与对 $(X_i-\mu_1)^2$ 和 $(Y_i-\mu_2)^2$ 的排序结果相同，但比它更容易.

假定条件

1. 两个样本都是来自各自总体的随机样本.
2. 除了每个样本内观测相互之间独立外，两个样本之间也相互独立.
3. 度量尺度至少是区间的.

检验统计量 如果 U 的值与 V 的值没有结，则赋给总体 1 的秩的平方和可以用作检验统计量.

$$T=\sum_{i=1}^{n}[R(U_i)]^2 \tag{3}$$

如果存在结，用 T 减去它的均值再除以它的标准差，就得到

$$T_1=\frac{T-n\overline{R^2}}{\left[\frac{nm}{N(N-1)}\sum_{i=1}^{N}R_i^4-\frac{nm}{N-1}(\overline{R^2})^2\right]^{\frac{1}{2}}} \tag{4}$$

其中，$N=n+m$，$\overline{R^2}$ 代表两个样本合并的平方秩的平均：

$$\overline{R^2}=\frac{1}{N}\left\{\sum_{i=1}^{n}[R(U_i)]^2+\sum_{j=1}^{m}[R(V_j)]^2\right\} \tag{5}$$

$\sum R_i^4$ 代表秩的四次幂的和：

$$\sum_{i=1}^{N}R_i^4=\sum_{j=1}^{n}[R(U_j)]^4+\sum_{j=1}^{m}[R(V_j)]^4 \tag{6}$$

零分布 当没有结，$n\leqslant 10$，$m\leqslant 10$ 时，表 A9 中给出了 T 的精确零分布的分位数. 当样本容量大于 10 时，表 A1 给出了下面基于标准正态分位数 z_p 的大样本逼近，它 301
可以用于获得 T 的近似分位数 w_p.

$$w_p = \frac{n(N+1)(2N+1)}{6} + z_p\sqrt{\frac{mn(N+1)(2N+1)(8N+11)}{180}} \tag{7}$$

其中，$N=n+m$.

T_1 的近似零分布是标准正态分布（见表 A1）

假设

A.（双边检验）

H_0: 除了它们的均值可能不同外，X 和 Y 同分布

H_1: $\mathrm{Var}(X) \neq \mathrm{Var}(Y)$

如果 T（或有结时的 T_1）大于它的 $1-\alpha/2$ 分位数或小于它的 $\alpha/2$ 分位数，则以显著性水平 α 拒绝 H_0，我们可以从表 A9 或（7）式得到 T 情况下的分位数，如果用 T_1，则分位数可以从表 A1 得到.

如果用 T_1，双边 p-值是 $P(Z \leqslant T_1)$ 或 $P(Z \geqslant T_1)$ 中较小者的 2 倍，这两个概率可以直接从表 A1 中获得. 如果用 T，可以从表 A9 中获得近似的 p-值，为找到导致拒绝 H_0 的最小双边检验，可用正态近似，

$$p\text{-值} = 2 \cdot (\text{单边}p\text{-值中较小的}) \tag{8}$$

其中，左边 p-值近似为：

$$\text{左边}p\text{-值} = P\left(Z \leqslant \frac{T - n(N+1)(2N+1)/6}{\sqrt{mn(N+1)(2N+1)(8N+11)/180}}\right) \tag{9}$$

而右边 p-值近似为：

$$\text{右边}p\text{-值} = P\left(Z \geqslant \frac{T - n(N+1)(2N+1)/6}{\sqrt{mn(N+1)(2N+1)(8N+11)/180}}\right) \tag{10}$$

B.（左边检验）

H_0: 除了它们的均值可能不同外，X 和 Y 同分布

302 H_1: $\mathrm{Var}(X) < \mathrm{Var}(Y)$

如果 T（或有结时的 T_1）小于它的 α 分位数，则以显著性水平 α 拒绝 H_0，我们可以从表 A9 或（7）式获得 T 情况下的分位数，如果用 T_1，则分位数可以从表 A1 得到. p-值是在零分布下小于或等于 T(或 T_1)的概率. 对于 T，它近似地由（9）式给出，而对于 T_1，用表 A1，由 $P(Z \leqslant T_1)$ 给出 p-值.

C.（右边检验）

H_0: 除了它们的均值可能不同外，X 和 Y 同分布

H_1: $\mathrm{Var}(X) > \mathrm{Var}(Y)$

如果 T（或有结时的 T_1）大于它的 $1-\alpha$ 分位数，则以显著性水平 α 拒绝 H_0，我们可以从表 A9 或（7）式获得到 T 情况下的分位数，如果用 T_1，则分位数可以从表 A1 得到. p-值是在零分布下大于或等于 T（或 T_1）观测值的概率. 对于 T，它近似地由（10）式给出，而对于 T_1，可直接从表 A1 得到 p-值.

计算机辅助 这个检验可以用 *StatXact* 实现，并称为 Conover 检验. ◄

多于两个样本的检验

如果有 3 个或更多的样本，这个检验可简单地修改为检验几个方差相等. 正像两样本时描述的那样，用每个观测减去它的总体均值（或当 μ_i 未知时，用它的样本均值），并将负值换为正值. 如前面两样本所述，对合并的绝对离差从小到大排序，当有结时，赋以平均秩. 计算每个样本秩的平方和，记 $S_1,S_2,\cdots,S_k$ 为 k 个样本的秩平方和. 这样，S_1 对应于前面两样本情形中的 T.

H_0:除了它们的均值可能不同外，所有k个总体同分布

H_1:有些总体的方差不彼此相等

检验统计量为：

$$T_2=\frac{1}{D^2}\left[\sum_{j=1}^{k}\frac{S_j^2}{n_j}-N(\bar{S})^2\right] \tag{11}$$

其中，n_j = 样本 j 中的观测数.

$$N=n_1+n_2+\cdots+n_k$$

$$S_j=\text{样本}\,j\,\text{中秩的平方和}$$

$$\bar{S}=\frac{1}{N}\sum_{j=1}^{k}S_j=\text{所有秩平方的平均}$$

$$D^2=\frac{1}{N-1}\left[\sum_{i=1}^{N}R_i^4-N(\bar{S})^2\right]$$

$\sum R_i^4$ 代表每个秩四次幂的和. 如果没有结，D^2 和 $\bar{S}$ 可简化为

$$D^2=N(N+1)(2N+1)(8N+11)/180 \tag{12}$$

和

$$\bar{S}=(N+1)(2N+1)/6 \tag{13}$$

303

零分布近似为自由度为 $k-1$ 的χ^2 分布，在表 A2 中给出了它的上分位数.

如果 T_2 超过从表 A2 得到的自由度为 $k-1$ 的χ^2 分布的 $1-\alpha$ 分位数，则拒绝零假设.

p-值近似为自由度为 $k-1$ 的χ^2 随机变量大于 T_2 的观测值的概率. 如果拒绝了 H_0，如前一节所述，我们可以作多重比较. 此时，如果下列不等式满足，就说总体 i 和 j 的方差不同.

$$\left|\frac{S_i}{n_i}-\frac{S_j}{n_j}\right|>t_{1-\alpha/2}\left(D^2\frac{N-1-T_2}{N-k}\right)^{\frac{1}{2}}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)^{\frac{1}{2}} \tag{14}$$

其中，$t_{1-\alpha/2}$是自由度为 $N-k$ 的 t 分布的 $1-\alpha/2$ 分位数，它可从表 A21 中获得.

例 5.3.1

食品包装公司想要相当肯定，它们生产的谷类食品包装盒里实际所含的谷类食品盎司数至少是包装盒外边贴的那个量. 为了做到这点，给每盒的平均量必须稍微超出它所做广告的量，因为包装机器可能会造成不可避免的变化量，给盒子装谷类食品有时会稍微多一点或稍微少一点. 因为每盒的平均量会调整到和广告上的平均量很接近，所以有较小变化量的机器会给公司省钱.

检验一台新机器，看它是否比现有的机器变化量更小，以便可以购买它来代替旧机器. 用现有的机器给几个盒子装满谷类食品，测量每个盒子装的量，对新机器也进行同样的操作，要检验

304

H_0:两台机器有相同的方差

H_1:新机器有较小的方差

测量值和计算结果如下.

原始测量		绝对离差		秩		秩平方	
现有的 (X)	新的 (Y)	现有的 (U)	新的 (V)	现有的	新的	现有的	新的
10.8	10.8	.06	.01	4	2 (结)	16	4
11.1	10.5	.36	.29	10	8	100	64
10.4	11.0	.34	.21	9	7	81	49
10.1	10.9	.64	.11	12	6	144	36
11.3	10.8	.56	.01	11	2 (结)	121	4
	10.7		.09		5		25
	10.8		.01		2 (结)		4
$\overline{X}$ = 10.74	$\overline{Y}$ = 10.79					T = 462	

$$T = \text{平方秩的和(现有的)} = 462$$

$$\overline{R^2} = \frac{1}{12}(16 + 100 + \cdots + 25 + 4) = 54$$

$$\sum_{i=1}^{N} R_i^4 = (16)^2 + (100)^2 + \cdots + (25)^2 + (4)^2 = 60{,}660$$

$$T_1 = \frac{462 - 5(54)}{\left[\frac{(5)(7)}{(12)(11)}(60{,}660) - \frac{(5)(7)}{11}(54)^2\right]^{\frac{1}{2}}} = 2.3273$$

前面的假设与检验集 C 相符，因为 H_1 指定新机器（Y）有更小的方差，则近似水平 α 为 0.05 的临界域对应着 T_1 的值大于 1.6449（表 A1 中的 0.95 分位数）. 在这种情况下，T_1 超过 1.6449，所以拒绝 H_0. 将观测到的 $T_1 = 2.3273$ 与表 A1 中的分位数比较，发现 p-值大约是 0.01.

在这个例子中，只要 U 的值和 V 的值没有结，计算结果就会有相当大的简化. 于是可以用秩而不用平均秩，并使用精确表. 在这个例子中，只有 V 的 3 个值存在结，所以当 3 个结的值出现时，在最右边的列中，用 $1^2=1, 2^2=4$ 和 $3^2=9$，而不用 $2^2=4$. 检验的其余部分按没有结来计算. 这次 T 的值碰巧没有变，对 $n=5, m=7$，它比表 A9 中的 0.95 分位数 410 大，而且表明这个近似检验的 p-值大约为 0.01 ■

□**理论** 只要两个随机变量 X 和 Y 除了有不同的均值 μ_1 和 μ_2 外是同分布的，则 $X-$
305 μ_1 和 $Y-\mu_2$ 不仅有零均值，而且它们也是同分布的. 这意味着 $U=|X-\mu_1|$ 和 $V=|Y-\mu_2|$ 有相同的分布，且 $U^2=(X-\mu_1)^2$ 和 $V^2=(Y-\mu_2)^2$ 也有相同的分布. 所以对

于 X 和 Y 的随机样本，U 和 V 是独立同分布的. 因此像 Mann-Whitney 检验中的那样，U 的秩的每种分配都是等可能的，且在 5.1 节中能找到关于秩的任何函数的分布.

注意，$U_1 < U_2$ 当且仅当 $U_1^2 < U_2^2$，所以 U 的秩与 U^2 相应的秩相同. 因为我们感兴趣的是比较 $E(U^2)$ 和 $E(V^2)$，故我们应该看 U^2 和 V^2 的秩；但 U 和 V 的秩是等价于 U^2 和 V^2 的秩而且更容易.

区别这个秩检验和前面秩检验的另外一个重要差异就是，所用的是平方秩而不是秩本身. 我们用的是得分而不是秩，用 R 的函数 $a(R)$ 来记得分，在检验统计量中用 $a(R)$ 来代替 R. 记 T 为关于一个样本的得分的和，在上述检验中，得分 $a(R) = R^2$. 我们可用像 5.1 节中的方法来求 T 的分布. 如果样本容量是 $n=3, m=4$，从 7 个秩中选出 3 个的方式有 35 种. 1，2，3 这 3 个秩相对应的得分是 $a(1), a(2)$ 和 $a(3)$，它们将用来计算检验统计量 T 的值（依赖于所用的得分），其概率为 1/35. 从 7 个秩中选出 3 个的 35 种方式给出了 35 个 T 值，可能有相同的，这时用 5.1 节中的方法很简单地就得到了 T 的概率函数.

当 H_0 为真时，对 T 用大样本的正态逼近，这时有必要求出 T 的均值和方差. 我们有

$$T = \sum_{i=1}^{n} a(R_i) \tag{15}$$

此时，$R_1, \cdots, R_n$ 代表 $U_1, \cdots, U_n$ 在 U 和 V 合并样本中的秩. 我们先对一般的得分 $a(R)$ 找出 $E(T)$ 和 $\mathrm{Var}(T)$，最后以 $a(R) = R^2$ 来代替.
由定理 1.4.1，T 的均值可写为：

$$E(T) = E\left[\sum_{i=1}^{n} a(R_i)\right] = \sum_{i=1}^{n} E[a(R_i)] \tag{16}$$

因为对每个 $j = 1, 2, \cdots, N, P(R_i = j) = 1/N$，我们有

$$E[a(R_i)] = \sum_{j=1}^{N} a(j) \cdot \frac{1}{N} = \frac{1}{N}\sum_{j=1}^{N} a(j) = \bar{a} \tag{17}$$

（记为 $\bar{a}$），对所有的 i 从 1 到 n，它都是相同的，所以（16）式变为

$$E(T) = n\bar{a} \tag{18}$$

其中，$\bar{a}$ 是所有得分的平均. 306

对于 T 的方差，由定理 1.4.3，我们得到

$$\mathrm{Var}\,(T) = \sum_{i=1}^{n} \mathrm{Var}\,[a(R_i)] + \sum_{\substack{i=1 \\ i\neq j}}^{n} \sum_{j=1}^{n} \mathrm{Cov}\,[a(R_i), a(R_j)] \tag{19}$$

其中，由方差的定义，

$$\mathrm{Var}\,[a(R_i)] = E\{[a(R_i) - \bar{a}]^2\} = \sum_{k=1}^{N} [a(k) - \bar{a}]^2 \cdot \frac{1}{N} = A \tag{20}$$

以及由协方差的定义，

$$\text{Cov}\,[a(R_i),a(R_j)] = E\{[a(R_i)-\bar{a}][a(R_j)-\bar{a}]\}$$
$$= \sum_{\substack{k=1 \\ k\neq l}}^{N}\sum_{l=1}^{N}[a(k)-\bar{a}]\,[a(l)-\bar{a}]\,\frac{1}{N(N-1)} \tag{21}$$

因为，对所有的 $k\neq l, P(R_i=k,R_j=l)=1/[N(N-1)]$，我们在（21）式的表达式中同时加上和减去 $k=l$ 的项，使之简化为：

$$\text{Cov}\,[a(R_i),a(R_j)] = \sum_{k=1}^{N}[a(k)-\bar{a}]\sum_{l=1}^{N}[a(l)-\bar{a}]\,\frac{1}{N(N-1)} - \sum_{k=1}^{N}[a(k)-\bar{a}]^2\,\frac{1}{N(N-1)} \tag{22}$$

但由于 $\bar{a}$ 在（17）式中已定义，故第一个和等于零，所以（22）式又简化为

$$\text{Cov}\,[a(R_i),a(R_j)] = -\frac{1}{N-1}A \tag{23}$$

其中，A 由（20）式定义. 现在将（20）式和（23）式中的方差和协方差项代入到（19）式，得到

$$\text{Var}\,(T) = \sum_{i=1}^{n}A - \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n}\frac{1}{N-1}A$$
$$= nA - n(n-1)\frac{1}{N-1}A$$
$$= \frac{n(N-n)}{N-1}A$$
$$= \frac{nm}{(N-1)N}\sum_{i=1}^{N}[a(i)-\bar{a}]^2 \tag{24}$$

由于 $N-n=m$. 当有结存在时，（18）式和（24）式在5.1节和5.2节中已用过，而
307 且在本章后面部分也会有用. 现在我们感兴趣的是平方秩检验中 $a(R)=R^2$ 的情形. 此时 $\bar{a}$ 可写为（4）式中 $\bar{R}^2$，用下面的恒等式：

$$\sum_{i=1}^{N}[a(i)-\bar{a}]^2 = \sum_{i=1}^{N}[a(i)]^2 - N(\bar{a})^2 \tag{25}$$

来简化计算，可知（4）式的分母是（24）式的平方根.

两样本情形推广到 k 个样本的情形完全类似于把两样本 Mann-Whitney 检验推广到 k 个样本的 Kruskal- Wallis 检验情形. 即对 k 个样本中的每一个找得分和，记为 $S_1,S_2,\cdots,S_k$，由（18）式和（24）式，S_i 的均值和方差为

$$E(S_i) = n_i\bar{a} \tag{26}$$

和

$$\text{Var}\,(S_i) = \frac{n_i\,(N-n_i)}{(N-1)N}\sum_{i=1}^{N}[a(i)-\bar{a}]^2 \tag{27}$$

如同在 Kruskal-Wallis 检验中那样，用 $(N-n_i)/N$ 乘以项 $[S_i-E(S_i)]^2/\text{Var}(S_i)$，$i=1,2,\cdots,k$，并加到一起得：

$$T_2 = \sum_{i=1}^{k} \frac{(S_i - n_i\bar{a})^2}{n_i D^2} \tag{28}$$

其中

$$D^2 = \frac{1}{N-1}\sum_{i=1}^{N}[a(i)-\bar{a}]^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N}[a(i)]^2 - N(\bar{a})^2\right\} \tag{29}$$

(28)式可简化为

$$T_2 = \frac{1}{D^2}\left[\sum_{j=1}^{k}\frac{S_j^2}{n_j} - N(\bar{a})^2\right] \tag{30}$$

当得分是秩的平方时，它与(11)式一致．建议使用的多重比较方法是一个近似的方法，但当样本容量变大后就成为精确方法． □

如果 X 和 Y 的总体服从正态分布，合适使用的统计量是两个“样本方差”的比值，

$$F = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}{\frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \bar{Y})^2} \tag{31}$$ 308

它服从 F 分布．在表 A22 中的 $k_1 = n-1$ 列，$k_2 = m-1$ 行给出了 F 的上侧分位数，下侧分位数没有给出．但可以由 $k_1 = m-1$ 列，$k_2 = n-1$ 行所得上侧分位数求倒数得到．合适的单边和双边检验也因此而得到．

正如 Siegel 和 Tukey(1960)所指出的，F 检验对正态分布的假设非常敏感．真正的分布可能是对称的，很像正态分布，例如双指数分布，但真正的显著性水平可能是假设显著性水平的 2 倍或 3 倍大．正因为如此，除非总体确实是正态的，F 检验不是一个很安全的检验．

当总体是正态分布时，如果用平方秩检验，而不是用 F 检验，则渐近相对效率(A. R. E.)仅有 $15/(2\pi^2) = 0.76$．然而，对于双指数分布，A. R. E. 是 1.08；对于均匀分布，A. R. E. 是 1.00．相同的效率也适用于 k 个样本的情形．因此，F 检验对正态分布假设的敏感性，加上它在一些常用非正态的情形下功效较低的特点，因此我们应尽量考虑非参数的方差检验．

在平方秩检验中，用 $\bar{X}$ 和 $\bar{Y}$ 代替 X 和 Y 真正的均值，使得检验是近似的而不是精确的，而检验统计量的精确分布依赖于总体的真实分布．正如 Conover，Johnson 和 Johnson(1981)对 56 个方差检验所做的广泛模拟研究所示，当总体分布相当偏斜时，检验会有问题．由于总体的偏斜，使得显著性水平可能变得很大，所以在这种情况下，我们推荐使用 X 和 Y 的样本中位数来调整代替样本均值．然而，这两种方法都给出了一个渐近分布自由的检验，即当样本容量变大时，近似方法将变为精确方法．

另一个受欢迎的两样本尺度问题的非参数检验是基于统计量

$$T=\sum_{i=1}^{n}\left[R(X_i)-\frac{n+m+1}{2}\right]^2 \tag{32}$$

其中，$P(X_i)$是X_i的秩，像Mann-Whitney检验中一样，它是由Mood(1954)提出的. Laubscher，Steffens和DeLange(1968)在相同分布函数的零假设下给出了精确表. Hollander(1963)给出了相关统计量

$$T=\sum_{i=1}^{n}[R(X_i)-\overline{R}_x]^2 \tag{33}$$

309 的零分布表，其中

$$\overline{R}_x=\frac{1}{n}\sum_{i=1}^{n}R(X_i)$$

Ansari和Bradley(1960)讨论了Mood检验和其他检验的A. R. E..

Conover和Iman(1978a)给出了平方秩检验的进一步讨论和更广泛的表. Talwar和Gentle(1977)检查了这个检验的微小变化. 其他尺度检验由Sen(1963),Puri(1965),Mielke(1967),Duran和Mielke(1968),Shorack(1969),Hwang和Klotz(1975)所考虑，Fligner和Killeen(1976)考虑了两样本的情况，Tsai，Duran和Lewis(1975)考虑了几个样本的情况. Conover,Johnson和Johnson(1981)对56个方差检验进行了全面比较. Lepage(1971,1973,1977),Mielke(1972),Duran,Tsai和Lewis(1976)提出了设计用来同时检测位置和尺度差异的检验. Gibbons(1967)和Hollander(1968)研究了尺度秩检验和位置秩检验的相关性. Moses(1963),van Eeden(1964),Basu和Woodworth(1967),Bauer(1972),Laubscher和Odeh(1976)以及Bhattacharayya(1977)进一步考虑了尺度参数估计. 如果位置参数未知，且可能不相等，一个修改的检验见Raghavachari(1965a),Puri(1968)和Nemenyi(1969). 进一步的参考文献可以在Duran(1976)的一个非常好的评论文章中找到，或者是在Daniel(1979)的参考文献中找到.

习题

1. 血库中心留有几个献血者心跳速率的记录.

男	58	76	82	74	79	65	74	86	
女	66	74	69	76	72	73	75	67	68

男士之间的变化显著地比女士之间的变化大吗？

2. 近几年来，大面积建起来了一个特定的水域，有住宅区发展，水坝等等. 将这个水域的水流速率的一个随机样本（每分钟立方英尺）与早些年的水流速率的样本作比较，看变化量是否改变.

现在的速率	32	36	41	27	35	48	31	28
之前的速率	39	21	58	46	30	22	17	19

310

方差有显著差异吗？

3. 将五年级的学生随机分配到3个不同的教室，来比较教学的3种不同的方法. 在年初测试一下每个学生达到的成绩水平（用一个标准化考试来测量），在年底再测试一次，每个学

生成绩的增长记录如下.

教学方法	成绩的增长
单元教学	0.7, 1.0, 2.0, 1.4, 0.5, 0.8, 1.0, 1.1, 1.9, 1.2, 1.5
个人学习	1.7, 2.1, −0.4, 0, 1.0, 1.1, 0.9, 2.3, 1.3, 0.4, 0.5
露天教室	0.9, 0.9, 1.0, 0, 0.1, −0.6, 2.2, −0.3, 0.6, 2.4, 2.5

这 3 种教学方法在方差上有差异吗？如果有，哪个方法在变化上有所不同？

4. 把一个投资班的学生分为 3 组，1 组教投资证券，第 2 组教投资蓝筹股，第 3 组教纯理论知识. 每个学生"投资"（仅在纸上）$ 10，000，并在 3 个月后评估假设的收益或损失，结果如下.

证券	蓝筹股	纯理论知识
146	176	−540
180	110	1052
192	212	642
185	108	−281
153	196	67

方差的差别显著吗？如果是，哪组是有显著的差异？

思考题

1. 对 $n=3, m=4$，求出（3）式中给出的 T 的精确分布，并将其与表 A9 给出的分位数作比较.
2. 证明（28）式与（30）式等价.
3. 对 $n=3, n=4$，求出由（32）式给出的 Mood 统计量以及由（33）式给出的 Hollander 统计量的精确分布. 311
4. 证明由（32）式所给出的 Mood 统计量的均值是 $n(N+1)(N-1)/12$.
5. Siegel 和 Tukey(1960) 给出了等方差的另一个检验. 对 X 和 Y 的合并样本进行排序，给最小的值赋以秩 1，给最大的值赋以秩 2，给第二大的值赋以秩 3，给第二小的值赋以秩 4，给第三小的值赋以秩 5，等等，交替地给两边的各两个值赋秩（在第一次后），一直进行到中间. 检验统计量是赋给样本 X 的秩和.
   (a) 说明当两个总体同分布时，表 A7 对统计量是有用的.
   (b) 对于单边备择假设 $\mathrm{Var}(X) > \mathrm{Var}(Y)$，应该用哪边（左或右）的临界域？
   (c) 用极端的例子来说明当两个总体的均值相差很远时，这个检验几乎没有功效.
6. 如果用得分 $a(i)$ 来计算（5.2.19）式中定义的 F 统计量，证明结果可以简化为如下形式.

$$F = \frac{T_2/(k-1)}{(N-1-T_2)/(N-k)}$$

其中，T_2 由（28）式和（30）式给出. 注意，这个数学关系式对所有类型的得分都成立.

5.4 秩相关性度量

相关性度量是一个用于由数对组成的数据情形中的随机变量，如二维数据. 假设一个容量为 n 的二维随机样本为 $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$，一般涉及 (X_i, Y_i) 时，我们将用 (X, Y). 即对 $i=1, \cdots, n$，(X_i, Y_i) 有相同的二维分布，并

和（X,Y）的二维分布相同.

二维随机变量的例子，如 X_i 代表第 i 个人的高度，Y_i 代表他父亲的高度，或者 X_i 代表第 i 个人的测试成绩，Y_i 代表她的训练量.

习惯上，可接受的 X 和 Y 之间的相关性度量应该满足下面的要求.

1. 相关性度量应该假设其值只在 -1 到 1 之间.

2. 如果 X 的大值倾向于和 Y 的大值配对，并且 X 的小值倾向于和 Y 的小值配对，则相关性度量应该是正的，如果趋势很强时，则接近于 +1.0. 那么，我们称 X 和 Y 之间正相关.

3. 如果 X 的大值倾向于和 Y 的小值配对，并且 X 的小值倾向于和 Y 的大值配
312 对，那么相关度量应该是负的，如果趋势很强时，应接近于 -1.0. 那么，我们称 X 和 Y 之间负相关.

4. 如果 X 的值和 Y 的值看上去是随机配对的，则相关性度量应该接近于零. 这大多数应该是 X 和 Y 独立时的情形，可能也有些 X 和 Y 并不独立的情形. 这时我们称 X 和 Y 不相关，或者是没有相关，或者有零相关.

最常用的相关性度量是 Pearson 乘积矩相关系数，记为 r，它定义为

$$r=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right]^{\frac{1}{2}}} \tag{1}$$

其中，$\bar{X}$ 和 $\bar{Y}$ 为样本均值，见 2.2 节所定义. 用于计算的较简单形式是

$$r=\frac{\sum_{i=1}^{n}X_iY_i-n\overline{X}\overline{Y}}{\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n}Y_i^2-n\overline{Y}^2\right)^{\frac{1}{2}}} \tag{2}$$

如果（1）式的分子和分母同除以 n，则 r 变为

$$r=\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]^{\frac{1}{2}}\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right]^{\frac{1}{2}}} \tag{3}$$

这可能容易记忆，因为样本协方差是分子，两个样本标准差的乘积是分母.

Pearson r 是 X 和 Y 之间线性关系强度的度量. 这意味着如果 Y 对 X 的散点图，显示点（X,Y）都落在或接近于一条直线时，那么，r 将等于或接近于 ±1.0，这里 +1.0 和 -1.0 视直线的斜率为正或负而定.

除了度量尺度至少是区间的外，Pearson r 比较难以解释. 尽管这样，这个相关性度量可用于任意的数值型数据，没有任何度量尺度或基础分布的要求，这就
313 迎合了可接受相关度量的必要要求. 然而，r 是随机变量，因而有分布函数. 不幸

的是，r 的分布依赖于 (X,Y) 的二维分布. 因此，除非 (X,Y) 的分布已知，作为非参数检验的检验统计量或者在构成置信区间时，r 没有太大的价值.

除了这个广泛被接受的 r 外，也发现了其他满足上述可接受要求的相关性度量. Kruskal(1958)的一篇非常优秀且可读的综述文章讨论了很多相关性度量. 如果 X 和 Y 是独立的，则一些相关性度量有分布函数，且它们不依赖于 (X, Y) 的二维分布函数，因此，它们可以用作非参数独立性检验的检验统计量. 这里选择表达的相关性度量是赋给观测值秩的函数. 如果 X 和 Y 是独立且连续的，则它们的分布函数和 (X,Y) 的二维分布函数是无关的. 如果数据是须序度量尺度的，则它们甚至可以用作某些非数值型数据的相关性度量. 我们提出的第一个秩相关系数是基于 X 和 Y 的秩计算的简单 Pearson r.

▶ Spearman ρ

数据 数据由样本容量为 n 的二维随机变量 $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$ 组成. 记 $R(X_i)$ 为 X_i 与其他 X 值相比的秩，$i=1,2,\cdots,n$. 即如果 X_i 是 $X_1,X_2,\cdots,X_n$ 中最小的，则 $R(X_i)=1$，如果 X_i 是 $X_1,X_2,\cdots,X_n$ 中第二小的，则 $R(X_i)=2$，等等，秩 n 被赋给最大的 X_i. 类似地，记 $R(Y_i)$ 等于 $1,2,\cdots$，或 n，它依赖于对每个 i, Y_i 与 $Y_1,Y_2,\cdots,Y_n$ 相比的相对大小.

或者，如果观测可用刚描述的方式排序，则数据可以由 n 对非数值的观测组成. 排序可以是基于观测的质量（"最差"观测或"最好"观测），或者根据对观测的喜好程度，等等.

像 Mann-Whitney 和 Kruskal-Wallis 检验一样，在有结存在的情况下，给每个有结的值赋以没有结时本应赋给它们的秩的平均值.

相关性度量 Spearman(1904)给出了相关性度量，经常记它为 ρ(rho)，定义如下： 314

$$\rho=\frac{\sum_{i=1}^{n}R(X_i)R(Y_i)-n\left(\frac{n+1}{2}\right)^2}{\left(\sum_{i=1}^{n}R(X_i)^2-n\left(\frac{n+1}{2}\right)^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n}R(Y_i)^2-n\left(\frac{n+1}{2}\right)^2\right)^{\frac{1}{2}}} \tag{4}$$

这是基于秩与平均秩计算的简单 Pearson r.

如果没有结，一个等价而计算简单的形式给出如下：

$$\rho=1-\frac{6\sum_{i=1}^{n}[R(X_i)-R(Y_i)]^2}{n(n^2-1)}=1-\frac{6T}{n(n^2-1)} \tag{5}$$

其中，T 代表整个分子的求和，这个形式仅在没有结时等价. 如果有很多结时，则用 (4) 式. 如果在数据中出现了中等个数的结，因为 (4) 式和 (5) 式之间的差别很小，则推荐使用计算简便的 (5) 式.

正如我们所说，Spearman ρ 仅仅是又将观测值换为它们的秩，然后基于秩用 Pearson r 计算得到的，这从下面的计算可以看到．如果数据用它们的秩代替，则 $\bar{X}$ 和 $\bar{Y}$ 对应于

$$\overline{R(X)} = \frac{1}{n}\sum_{i=1}^{n} R(X_i) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}\frac{n(n+1)}{2}$$
$$= \frac{n+1}{2} \tag{6}$$

和

$$\overline{R(Y)} = \frac{n+1}{2} \tag{7}$$

所以（2）式变为（4）式.

例 5.4.1

测量12个MBA研究生的入学考试GMAT分数和他们读MBA项目时的平均成绩（GPA），以研究GMAT和GPA分数之间的关系强度．他们的GMAT成绩和GPA

315 如下给出，并附带有秩和一些计算.

学生	GMAT(X)	GPA (Y)	$R(X)$	$R(Y)$	$[R(X)-R(Y)]^2$
1	710	4.0	12	11.5	0.25
2	610	4.0	9.5	11.5	4
3	640	3.9	11	10	1
4	580	3.8	8	9	1
5	545	3.7	3	8	25
6	560	3.6	5	7	4
7	610	3.5	9.5	5	20.25
8	530	3.5	1	5	16
9	560	3.5	5	5	0
10	540	3.3	2	3	1
11	570	3.2	7	1.5	30.25
12	560	3.2	5	1.5	12.25

因为有一半的观测都有结，所以应该用（4）式.

$$\sum_{i=1}^{12}[R(X_i)]^2 = 647.5 \qquad \sum_{i=1}^{12}[R(Y_i)]^2 = 647$$

由引理1.4.2，在没有结时，它们等于 $n(n+1)(2n+1)/6=650$．而且

$$\sum_{i=1}^{12} R(X_i)R(Y_i) = 589.75 \tag{8}$$

将上面的结果代入（4）式，有

$$\rho = \frac{589.75 - 12\left(\frac{13}{2}\right)^2}{\left(647.5 - 12\left(\frac{13}{2}\right)^2\right)^{1/2}\left(647 - 12\left(\frac{13}{2}\right)^2\right)^{1/2}} = 0.5900 \tag{9}$$

为了比较，(5) 式给出

$$\rho = 1 - \frac{6(115)}{12(12^2 - 1)} = 0.5979 \tag{10}$$

它比精确的 $\rho = 0.5900$ 稍微大一点，差别是因为在有结时用了平均秩.

有兴趣的一点是基于原始数据计算的 Pearson r 是 $r = 0.6630$，在这种情况下，X 和 Y 之间显然出了比 X 的秩和 Y 的秩之间更强的线性关系. ■

假设检验 Spearman 秩相关系数经常用作两个随机变量之间独立性检验的检验统计量，见上面给出的 Spearman ρ 的数据一节，(4) 式给出了检验统计量. 316

零分布 当 X 和 Y 独立，$n \leqslant 30$，且没有结时，ρ 的精确分位数可由表 A10 给出. 对于较大的 n，或者有很多结，ρ 的 p-分位数近似由下式给出

$$w_p = \frac{z_p}{\sqrt{n-1}} \tag{11}$$

其中，z_p 可在表 A1 中查到，它是标准正态分位数.

假设 Spearman ρ 对某些类型的相关不敏感，所以最好明确我们所要检测哪种类型的相关. 因此，假设有以下形式：

A. （双边检验）

H_0: X_i 和 Y_i 互相独立

H_1: 或者 (a) 较大的 X 值倾向与较大的 Y 值配对，或者 (b) 较小的 X 值倾向与较大的 Y 值配对

如果 ρ 的绝对值 $|\rho|$，大于它的 $1-\alpha/2$ 分位数（分位数可以从表 A10 或 (11) 式得到），则以水平 α 拒绝 H_0，且近似的双边 p-值是（使用表 A1）：

$$p\text{-值} = 2 \cdot P(Z \geqslant |\rho| \sqrt{n-1}) \tag{12}$$

B. （负相关左边检验）

H_0: X_i 和 Y_i 互相独立

H_1: 较小的 X 值倾向与较大的 Y 值配对，并且较大的 X 值倾向与较小的 Y 值配对

如果 $\rho < -w_{1-\alpha}$（$w_{1-\alpha}$ 可以从表 A10 或 (11) 式得到），则以水平 α 拒绝 H_0. 近似的左边 p-值是（使用表 A1）：

$$p\text{-值} = P(Z \leqslant \rho \sqrt{n-1}) \tag{13}$$

C. （正相关右边检验）

H_0: X_i 和 Y_i 互相独立

H_1: 较大的 X 的值倾向与较大的 Y 值配对 317

如果 $\rho > w_{1-\alpha}$（$w_{1-\alpha}$ 可以从表 A10 或 (11) 式得到），则以水平 α 拒绝 H_0. 近似的右边 p-值是（使用表 A1）：

$$p\text{-值} = P(Z \geqslant \rho \sqrt{n-1}) \tag{14}$$

计算机辅助 *Minitab*, *S-Plus*, *SAS* 和 *StatXact* 可以计算 Spearman ρ，并作独立性检验. 这些和其他程序都会将数据转换为秩，并用秩计算 Pearson r，在有结存在时，它会

自动校正.

例5.4.2

让我们继续例5.4.1. 假设例5.4.1中的12个MBA研究生是所有近期MBA研究生的随机样本，我们想知道是否有GPA高的学生，他的GMAT成绩也高的趋势. 则零假设是

$$H_0\text{: GPA和GMAT成绩互相独立}$$

感兴趣的备择假设是

$$H_1\text{: GPA高的学生, 他的GMAT成绩也高}$$

这和右边检验的假设C相符合. 因此，如果ρ超过它的0.95分位数，就以水平$\alpha=0.05$拒绝H_0. 由于有大量的结存在，则在表A10中，$n=12$的分位数$w_{0.95}=0.4965$只是近似值. 正态近似值

$$w_{0.95}=\frac{1.6449}{\sqrt{11}}=0.4960$$

也几乎是一样的，但可能较精确.

观测到的ρ值是0.5900，所以我们可以安全的断言，近期的MBA研究生的GPA和GMAT成绩之间有正相关. 从（14）式，p-值可近似为：

$$p\text{-值}=P(Z\geqslant 0.5900\sqrt{11})=P(Z\geqslant 1.9568)=0.025$$

■

接下来我们要提出与Spearman ρ类似的相关性度量，它是基于观测的排序（秩）而不是数值本身. 如果X和Y独立且连续时，这个度量的分布将不依赖于X和Y的分布. 这个度量称为Kendall τ（tau），经常认为它比Spearman ρ更难计
318 算. Kendall τ的主要优点就是它的分布能非常快地接近于正态分布，使得当X和Y独立的零假设为真时，Kendall τ的正态逼近比Spearman ρ的要好. Kendall τ的另一个优点是它的解释直接而简单，可根据观测协调和不协调的对的概率来解释，它们将在下面定义.

▶Kendall τ

数据 数据可以由一个容量为n的二维随机样本(X_i, Y_i)，$i=1, 2, \cdots, n$组成. 两个观测称做是协调的（concordant），如果一个观测的两个元素都比它们对应的另一观测的元素大，例如（1.3,2.2）和（1.6,2.7）. 记N_c为协调观测的对数，它是总共$\binom{n}{2}$可能对的一部分. 一对观测称做是不协调的（discordant），如果一个观测的两个数与它们对应的另一观测的数大小反向（相应的差值一个为正，一个为负），例如（1.3,2.2）和（1.6,1.1）. 记N_d为不协调观测的对数. 元素之间各自有结的对

的情形，按下面在有结一节讨论的来计数. 由于 n 个观测可能有 $\binom{n}{2}=n(n-1)/2$ 种不同的方式配对，协调对个数 N_c，不协调对个数 N_d，与带有结的对数之和将等于 $n(n-1)/2$.

如果观测可以使得刚描述的 N_c 和 N_d 能计算，则数据也可以由 n 对非数值观测组成.

相关性度量 Kendall（1938）提出的没有结的相关性度量如下：

$$\tau=\frac{N_c-N_d}{n(n-1)/2} \tag{15}$$

如果所有的对都是协调的，则 Kendall τ 等于 1.0. 如果所有的对都是不协调的，则值为 -1.0. 作为相关性度量，Kendall τ 满足本节一开始的要求.

结 用更精确的语言描述，如果 $(Y_2-Y_1)/(X_2-X_1)$ 大于 0，则一对二维观测 (X_1,Y_1) 和 (X_2,Y_2) 认为是协调的；如果它小于 0，则认为是不协调的. 如果 $X_1=X_2$，分母为 0，所以不作比较. 然而，如果 $Y_1=Y_2$（且 $X_1\neq X_2$），比值 $(Y_2-Y_1)/(X_2-X_1)$ 是 0，在这种情形下，这个数对应该计数为 1/2 协调和 1/2 不协调. 这时 τ 的分子并没有什么差异，因为当计算 N_c-N_d 时 1/2 项相抵消. 然而，它使得有结时，计算 τ 的方式不同了. 319

在有结时，我们可以用

$$\tau=\frac{N_c-N_d}{N_c+N_d} \tag{16}$$

这里将所有 $X_i\neq X_j$ 的对 (X_i,Y_i) 和 (X_j,Y_j) 进行比较. 这一形式的 Kendall τ 有这样一个优点，即使有结时也可以得到 +1 或 -1. Goodman 和 Kruskal（1963）首先对它进行了讨论，有时也把它叫做 γ 系数（gamma coefficient）.

总的来说，

如果 $\frac{Y_j-Y_i}{X_j-X_i}>0$，给 N_c 加 1（协调）

如果 $\frac{Y_j-Y_i}{X_j-X_i}<0$，给 N_d 加 1（不协调）

如果 $\frac{Y_j-Y_i}{X_j-X_i}=0$，给 N_c 和 N_d 各加 1/2

如果 $X_i=X_j$，不作比较.

如果根据 X 值的增大将观测 (X_i,Y_i) 排在一列上，则 τ 的计算就会被简化，那么每个 Y 就只与它下面的值进行比较，且协调与不协调的个数也就容易确定了，而且每对观测只考虑一次. 这个方法在下面的例子中给予解释.

例 5.4.3

我们还利用例 5.4.1 中的数据来解释. 根据 X 值的增大将数据 (X_i, Y_i) 排列如下.

	X_i, Y_i	协调对 (X_i, Y_i) 之下	不协调对 (X_i, Y_i) 之下
	(530, 3.5)	7	4
	(540, 3.3)	8	2
	(545, 3.7)	4	5
结	(560, 3.2)	5.5	0.5
结	(560, 3.5)	4.5	1.5
结	(560, 3.6)	4	2
	(570, 3.2)	5	0
	(580, 3.8)	3	1
结	(610, 3.5)	2	0
结	(610, 4.0)	0.5	1.5
	(640, 3.9)	1	0
	(740, 4.0)		
		$N_c = 44.5$	$N_d = 17.5$

320

Kendall τ 由下式给出

$$\tau = \frac{N_c - N_d}{N_c + N_d} = \frac{44.5 - 17.5}{44.5 + 17.5} = 0.4355$$

如 Kendall τ 所度量的，GMAT 成绩和 GPA 之间有正的秩相关. ■

假设检验 Kendall τ 也可以用做检验 X 和 Y 相互独立的零假设检验统计量，如同 Spearman ρ 所描述的那样，使用可能的单边或双边备择. 一些算法或许保留，然而，我们可直接用 $N_c - N_d$ 而不需要除以 $n(n-1)/2$ 作为检验统计量来获得 τ，因此用 T 作为 Kendall 检验统计量，这里 T 定义为

$$T = N_c - N_d \tag{17}$$

见上面给出的 Kendall τ 数据一节. 在没有结或结很少时，用（17）式给出的检验统计量给出；如果结很多，则应该用（16）式给出 τ.

零分布 当 X 和 Y 独立，$n \leqslant 60$，且没有结时，τ 和 T 的精确上侧分位数由表 A11 给出. 下侧分位数是这个表给出的上分位数的负数. 对于较大的 n，或者有很多结，τ 的 p 分位数近似表示如下：

$$w_p = z_p \frac{\sqrt{2(2n+5)}}{3\sqrt{n(n-1)}} \tag{18}$$

其中，z_p 是标准正态随机变量的 p 分位数（在表 A1 中给出）. T 的 p 分位数近似表示如下：

$$w_p = z_p \sqrt{n(n-1)(2n+5)/18} \tag{19}$$

假设

A.（双边检验）

H_0: X 和 Y 独立

H_1: 观测对倾向于或者协调，或者不协调

如果 T（或 τ）小于它的零分布的 $\alpha/2$ 分位数或大于它的 $1-\alpha/2$ 分位数（见表 A11），则以水平 α 拒绝 H_0. 321

双边 p-值是 2 倍的单边 p-值中较小者，近似表示如下

$$p(\text{左边}) = P\left(Z \leqslant \frac{(T+1)\sqrt{18}}{\sqrt{n(n-1)(2n+5)}}\right) \tag{20}$$

和

$$p(\text{右边}) = P\left(Z \geqslant \frac{(T-1)\sqrt{18}}{\sqrt{n(n-1)(2n+5)}}\right) \tag{21}$$

其中，T 是 $T_c - T_d$ 的观测值，连续相关是 1，Z 是标准正态随机变量，它的概率由表 A1 给出.

B.（左边检验）

H_0: X 和 Y 独立

H_1: 观测对倾向于不协调

如果 T（或 τ）小于它的零分布的 α 分位数（见表 A11），则以水平 α 拒绝 H_0. 左边 p-值近似地由（20）式给出.

C.（右边检验）

H_0: X 和 Y 独立

H_1: 观测对倾向于协调

如果 T（或 τ）大于它的零分布的 $1-\alpha$ 分位数（见表 A11），则以水平 α 拒绝 H_0. 右边 p-值近似地由（21）式给出.

计算机辅助 *Minitab*, *S-Plus*, *SAS* 和 *StatXact* 可以计算 Kendall τ，并可以作独立性检验.

例 5.4.4

在例 5.4.3 中 Kendall τ 由先求出下列 T 值来计算

$$T = N_c - N_d = 44.5 - 17.5 = 27$$

如果我们感兴趣的是用 T 来检验零假设：学生的 GMAT 成绩和他或她的 GPA 独立，看高的 GPA 是否与高的 GMAT 成绩相关，那么如果 T 大于 $w_{0.95}=24$（在表 A11 中获得），则以水平 $\alpha=0.05$ 拒绝零假设. 因为 $T=27$，则拒绝零假设. 右边 p-值从（21）式近似得到. 322

$$\begin{aligned} p\text{-值} &= P(T \geqslant 27) \\ &\cong P\left(Z \geqslant \frac{(27-1)\sqrt{18}}{\sqrt{12\cdot 11\cdot 29}}\right) \\ &= P(Z \geqslant 1.7829) \\ &= 0.037 \end{aligned}$$

如果我们用（16）式给出的 τ 作为检验统计量，由于有结，结果是类似的. ■

同样的数据用于 Spearman ρ 和 Kendall τ 是为了更好地比较这两个统计量，可以看出，Spearman $\rho(\rho=0.5900)$ 比 Kendall $\tau(\tau=0.4355)$ 大. 然而，使用这两个统计

量的那两个检验（或它们的等价统计量）得出几乎同样的结果. 前面的两个陈述在大多数情形下都为真，但不是全部情形. Spearman ρ 的绝对值比 Kendall τ 大. 然而，作为显著性的检验，由于在很多情形下两个产生几乎相同的结果，所以没有一个很强的理由让我们喜欢一个而不喜欢另一个.

▶Daniels 趋势性检验

Daniels(1950)通过将度量（称为 X_i）与取度量的时间（或次序）配对，提出用 Spearman ρ 作趋势性检验. 假设 X_i 相互独立，零假设是它们同分布，备择假设是 X_i 的分布与时间有关，使得当时间增加时，X 的度量趋向于变大（或变小）. 在 3.5 节中我们较全面地讨论了趋势性的概念，提出了 Cox 和 Stuart 趋势性检验. 一般认为基于 Spearman ρ 和 Kendall τ 的趋势性检验比 Cox 和 Stuart 的检验有更好的功效. 根据 Stuart(1956)，当应用到正态分布的随机变量，且基于回归系数检验时，3.5 节中所提到的 Cox 和 Stuart 趋势性检验的渐近相对效率（A. R. E.）大约是 0.78，而在同样的条件下，用 Spearman ρ 和 Kendall τ 检验的 A. R. E. 是 0.98. 但是，这些检验并不像 Cox 和 Stuart 检验那样可广泛应用. 例如，在例 3.5.3 中它们就不适用，而例 3.5.2 中则适用. 所以，我们用这个例子来解释 Spearman ρ 的趋势性检验. 使用 Kendall τ 检验的过程是类似的. ◀

例 5.4.5

323 在例 3.5.2 中，给出了 19 年的年降水量记录. 趋势性的双边检验包括

年 X_i	降水量 Y_i (英寸)	$R(X_i)$	$R(Y_i)$	$[R(X_i)-R(Y_i)]^2$
1950	45.25	1	12	121
1951	45.83	2	15	169
1952	41.77	3	11	64
1953	36.26	4	6	4
1954	45.27	5	13	64
1955	52.25	6	17	121
1956	35.37	7	2.5	20.25
1957	57.16	8	18	100
1958	35.37	9	2.5	42.25
1959	58.32	10	19	81
1960	41.05	11	9	4
1961	33.72	12	1	121
1962	45.73	13	14	1
1963	37.90	14	7	49
1964	41.72	15	10	25
1965	36.07	16	4	144
1966	49.83	17	16	1
1967	36.24	18	5	169
1968	39.90	19	8	121
				总和 1421.5

如果 Spearman ρ 太大或太小，则拒绝没有趋势的零假设. 因为结的个数不多，所以检验统计量由（5）式给出.

$$T=\sum_{i=1}^{19}[R(X_i)-R(Y_i)]^2=1421.5$$

$$\rho=1-\frac{6T}{19(19^2-1)}=1-\frac{6(1421.5)}{6840}=-0.2469$$

对 $\alpha=0.05$，ρ 的分位数（$n=19$，由表 A10 给出）为，

$$w_{0.975}=0.4579$$

和

$$w_{0.025}=-0.4579$$

像以前一样，接受 H_0.（12）式给出近似的 p-值：

$$\begin{aligned}p\text{-值}&\cong 2\cdot P(Z\geqslant 0.2469\sqrt{19-1})=2\cdot P(Z\geqslant 1.0475)\\&=2(0.147)\\&=0.294\end{aligned}$$

■ 324

▶Jonckheere-Terpstra 检验

Spearman ρ 或者 Kendall τ，可以用于几个独立样本的情形来检验零假设：所有的样本来自相同的分布，即

$$H_0: F_1(x)=F_2(x)=\cdots=F_k(x)$$

有序的备择假设：分布在指定的有序方向上

$$H_1: F_1(x)\geqslant F_2(x)\geqslant\cdots\geqslant F_k(x)$$

至少有一个不等式成立. 这个备择有时也写为

$$H_1: E(Y_1)\leqslant E(Y_2)\leqslant\cdots\leqslant E(Y_k)$$

其中，Y_i 代表分布函数为 $F_i(x)$ 的随机变量. 注意，这里的数据集和零假设都与 5.2 节中 Kruskal-Wallis 检验相同. 然而，Kruskal-Wallis 检验对均值的任何（any）差异都敏感，而 Spearman ρ，或者 Kendall τ 仅对上面给出 H_1 中的特殊有序敏感. 当用 Kendall τ 时，这个检验和 Jonckheere-Terpstra 检验等价，这可以在 *SAS* 和 *StatXact* 计算机程序中找到. 我们将在下面例 5.4.6 中解释这个程序. ◀

例 5.4.6

当人年纪增大，眼睛会看不清楚近处的物体，这是公认的 40 岁以上人的特征. 为了看 15～30 岁范围的人随着年龄的增长是否也失去了聚焦近物的能力，从 4 个年龄组的每一组中选择了 8 个人；这 4 个年龄组分别为：15 岁左右，20 岁左右，25 岁左右和 30 岁左右. 假设这些人对测量的特征来说是来自各自年龄组总体的随机样本. 每个人拿一张印有字的纸放在右眼前，左眼被遮住，把纸移近眼睛直到这个人称纸上的字变得模糊了，对每个人，测量纸上的字仍然清晰的最接近眼睛的距离.

零假设是对所有的总体测量的距离是同分布的，备择假设是年龄大的组的测量距离趋向于比较大．根据年龄组对样本从1到4编号．

$$H_0: F_1(x) = F_2(x) = F_3(x) = F_4(x) \qquad \text{对所有的}x$$

325 $$H_1: F_i(x) > F_j(x) \qquad \text{对某些}x\text{和某些}i < j$$

我们假设聚焦近物的能力没有随年龄提高，因此我们可以用稍简单的形式陈述零假设．

接下来，记测量距离为 Y（以英寸为单位），为了方便，样本自身做了排序，样本序号为 X．

15岁		20岁		25岁		30岁	
X	Y	X	Y	X	Y	X	Y
1	4.6	2	4.7	3	5.6	4	6.0
1	4.9	2	5.0	3	5.9	4	6.8
1	5.0	2	5.1	3	6.6	4	8.1
1	5.7	2	5.8	3	6.7	4	8.4
1	6.3	2	6.4	3	6.8	4	8.6
1	6.8	2	6.6	3	7.4	4	8.9
1	7.4	2	7.1	3	8.3	4	9.8
1	7.9	2	8.3	3	9.6	4	11.5

注意，如果最小聚焦距离 Y 的值随着年龄增大而增大，那么 Y 和 X 正相关，可使用 Spearman ρ，或者 Kendall τ，因此我们用 X 有很多结时的右边检验．同时注意到，我们可以用任意数值的增序列来替换 $X=1,2,3$ 和4以代表年龄组，如 $X=15,20,25$ 和30．更换后的 X 值不会改变 ρ 和 τ 的值．

这些数据的 Spearman ρ（我们省略计算细节）是 $\rho=0.5680$．5%右边检验的近似0.95分位数是 $1.6449/\sqrt{31}=0.2954$，有利于有序的备择，所以很容易拒绝零假设．右边 p-值小于0.001．

基于 $N_c=290.5, N_d=93.5$，这些数据的 Kendall τ 是 $\tau=0.5130$（仍然省掉细节）．与 $w_{0.95}=0.2056$ 比较，再次表明以 $\alpha=0.05$ 容易拒绝 H_0．右边 p-值仍小于0.001．这个检验用 Kendall τ 与 Jonckheere（1954a）和 Terpstra（1952）介绍的方法是等价的，尽管 Jonckheere 的检验统计量只是简单的 N_c（协调观测对的个数）．■

□**理论**　原则上，ρ 和 τ 的精确分布很容易得到，尽管在实际中甚至对中等的样本容量 n，其过程相当冗长乏味．精确分布是在 X_i 和 Y_i 独立同分布的假设下求出的，那么 $n!$ 个 X_i 的秩与 Y_i 的秩配对的排列是等可能的．如本章的前几节，分布函数很容易得到，因为通过计算给出 ρ 和 τ 为指定值时这种排列的个数，然后用这个数除以 $n!$ 就可得到 ρ 和 τ 为指定值的概率．

因为 ρ 与 τ 都是基于随机变量的和，所以可应用中心极限定理得到大样本的近
326 似分布，ρ 与 τ 的概率分布关于零对称，所以它们两个的均值都是零，而方差比较难得到，这里不打算推导．当 n 很大时，ρ 与 τ 除以各自的标准差得到的随机变量将近似于标准正态随机变量．对 $n\geqslant 8$，认为这个近似对求 τ 的分位数是相当好的，而在

求 ρ 的分位数时就不见得好了. □

如果$(X_i, Y_i), i=1,2,\cdots,n$ 是独立同分布的二维正态分布随机变量，相对于用 Pearson r 作为检验统计量的参数独立性检验（Stuart,1954），ρ 与 τ 都有渐近相对效率 $9/\pi^2 = 0.912$.

▶Kendall 偏相关系数

偏相关的概念不太容易理解，但为了解释可以把 Kendall τ 推广到偏相关的方式，我们将试图简要地描述一下偏相关

对多元随机变量（$X_1, X_2, \cdots, X_k$），在 X_1 和 X_2 之间，X_2 和 X_5 之间等等都可能有相关，这个相关的度量可以是已经描述过的任何一个度量. 这些度量估计的是一个随机变量对另一个的总影响（相关），包括有间接的影响，因为第二个随机变量不只与第一个随机变量相关，而且可能会和第三个随机变量有关，而第三个也可能又会与第一个随机变量有关，因此它会在第一个和第二个随机变量之间传递间接影响.

有时，我们要在以某种方式消除其他随机变量引起间接影响的条件下，来度量两个随机变量之间的相关. 当消除了由 $X_3, X_4, \cdots, X_n$ 引起的间接影响后，作为 X_1 和 X_2 之间相关的估计，称为"偏"相关估计，当用 Pearson r 进行推广时，它被记为 $r_{12.34\cdots n}$，当用 Kendall τ 进行推广时，记它为 $\tau_{12.34\cdots n}$.

在 $n=3$ 的简单情形下，偏相关可以用 Pearson 偏相关系数估计

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \tag{22}$$

其中，r_{ij}是通常在 X_i 和 X_j 之间计算的 Pearson r，也可以用 Kendall 偏相关系数估计

$$\tau_{12.3} = \frac{\tau_{12} - \tau_{13}\tau_{23}}{\sqrt{(1-\tau_{13}^2)(1-\tau_{23}^2)}} \tag{23}$$ 327

其中，τ_{ij}是通常在 X_i 和 X_j 之间计算的 Kendall τ. 可以用 *Minitab* 计算这些偏相关系数. ◀

Bartels(1982), Chan 和 Tran(1992), Cox(1966), Dufour(1981), Dufour 和 Roy(1985), Hallin 和 Melard(1988), Hallim et al.(1985), Hannan(1976), Harel 和 Puri(1990), Knoke(1977), Rao(1993), Sen(1981) 以及 Tran(1990) 讨论过用秩相关方法（通过用秩序列相关系数）检验一系列观测的相关性. 相关性的秩检验比另一受欢迎的非参数游程检验有更大的功效.

与 Kendall τ 中描述的一样，Spearman ρ 也被推广来度量偏相关. 用 Spearman ρ 推广的一个优点是求 Pearson 偏相关系数的已有计算机程序都可以用秩来计算，而不用数据，因此秩相关系数很容易得到.

$r_{12.3}$的分布依赖于（X_1, X_2, X_3）的多元分布函数，因此可能不会用作非参数检验的检验统计量. $\tau_{12.3}$和 $\rho_{12.3}$的分布也依赖于多元分布，因此分布不是自由的，除了所有 3 个变量相互独立之外. 这个主题的更多内容可参见 Simon(1977a, 1977b), Agresti

(1977)，或 Wolfe(1977b). Kendall(1942)给出了偏秩相关的讨论.

对用于另一情形时，Kendall 提出了另一个相关性度量是协调系数（coefficient of concordance). 当涉及更多的变量时，它可以用于度量总相关. 然而，Kendall 协调系数与 Friedman 提出的检验统计量之间有着密切的关系，这建议我们同时使用这些统计量，5.8 节中将详细讨论它

在 Kendall 和 Gibbons(1980)的书中包含了秩相关的广泛研究，也可参见 Gibbons (1993)，它包括了一些 *Minitab* 和 *SPSS* 的例子. Knight(1966)给出一个计算 Kendall τ 的计算机方法. Best(1973,1974)提出 Kendall τ 的推广表格，甚至有 $n \leqslant 25$，并且在有结时不同情形下的表. Stuart(1963)解释了列联表的 Spearman ρ. Zar(1972)给出了 Spearman ρ 更广泛的表，他用的是一些近似的方法，效果也相当不错. Iman 和 Conover(1978)比较了几种近似方法. Spearman ρ 的动态解释由 Evans(1973)给出

在回归分析中，Hotelling 和 Pabst(1936), Konijn(1961), Adichie(1967a,1967b), 和 Sen(1968a)讨论过秩相关方法的使用，它也是本章下两节的主要内容. 其他有关秩相关和相关性概念的文章有 Aitkin 和 Hume(1965), Lehmann(1966), Bell 和 Doksum (1967), Gokhale(1968), Ruymgaart et al. (1972), Ruymgaart(1973), Choi(1973)以及
328 Shirehata（1975，1976). Daniel(1980）的参考文献中列出了更多的文献.

习题

1. 小俩口一起去打保龄球，将他们所得的分数列为 10 行，看他们之间的成绩是否相关. 分数是：

行号	丈夫的分数	妻子的分数	行号	丈夫的分数	妻子的分数
1	147	122	6	151	120
2	158	128	7	196	108
3	131	125	8	129	143
4	142	123	9	155	124
5	183	115	10	158	123

(a) 计算 ρ .

(b) 计算 τ.

(c) 用基于 ρ 的双边检验来检验独立性假设.

(d) 有 τ 来做（c）中的问题.

2. 下面是 τ 与 ρ 产生变化很大的相关估计的一个例子.

X_i	Y_i	X_i	Y_i	X_i	Y_i
−8.7	−0.6	−1.9	−4.7	2.2	3.8
−8.3	−0.8	−1.6	−5.5	4.0	3.5
−8.2	−1.3	−1.3	−5.6	5.6	3.1
−7.2	−1.9	−0.2	−6.0	5.9	2.6
−6.1	−2.0	0.7	4.6	6.2	2.0
−6.0	−2.1	1.3	4.4	6.6	1.2
−4.1	−4.0	1.6	4.2	6.7	0.6
−2.0	−4.6	2.1	3.9	8.1	0.4

(a) 作一个粗略的散布图.

(b) 计算 τ.

(c) 计算 ρ.

(d) ρ 或 τ 会导致拒绝 X 和 Y 独立的零假设吗?

3. 分配一位新工人操作一台机器生产门闩. 每天抽查门闩的一个样本, 并记录次品率. 下面的数据说明这个工人随时间有显著的进步吗?

天	百分比	天	百分比	天	百分比
1	6.1	6	6.1	10	4.6
2	7.5	7	5.3	11	3.0
3	7.7	8	4.5	12	4.0
4	5.9	9	4.9	13	3.7
5	5.2				

(a) 用 Spearman ρ.

(b) 用 Kendall τ. 329

4. 美国总统第一次举行就职典礼的年龄和他去世的年龄有显著相关性吗?

姓名	就职年龄	去世年龄	姓名	就职年龄	去世年龄
Washington	57	67	Hayes	54	70
J. Adams	61	90	Garfield	49	49
Jefferson	57	83	Arthur	50	56
Madison	57	85	Cleveland	47	71
Monroe	58	73	Harrison	55	67
J.Q. Adams	57	80	McKinley	54	58
Jackson	61	78	T. Roosevelt	42	60
Van Buren	54	79	Taft	51	72
Harrison	68	68	Wilson	56	67
Tyler	51	71	Harding	55	57
Polk	49	53	Coolidge	51	60
Taylor	64	65	Hoover	54	90
Fillmore	50	74	F. Roosevelt	51	63
Pierce	48	64	Truman	60	88
Buchanan	65	77	Eisenhower	62	78
Lincoln	52	56	Kennedy	43	46
A. Johnson	56	66	L. Johnson	55	64
Grant	46	63	Nixon	56	81

(a) 用 Spearman ρ.

(b) 用 Kendall τ.

注意, 这些数据并不代表随机样本, 但可以假设他们是所有过去、现在、未来美国总统的随机样本

5. 5 名博士研究生参加了一次时事考试, 博士生的年龄和考试成绩如下.

年龄	24	31	38	45	45
考试成绩	68	85	84	92	90

年龄大学生的考试成绩会更高吗?

(a) 用 Spearman ρ.

(b) 用 Kendall τ.

6. 为了看最后一次课和期末考试时间之间留出更多时间是否会趋向于提高学生在期末考试中的成绩，将48名学生随机分为4组，每组12个学生. 第1组在最后一次课的2天后考试；第2组在最后一次课的4天后考试；第3组给6天时间；第4组给8天时间. 所有组
330 都给了在不同可比条件下的可比较考试. 期末考试成绩如下.

第1组			第2组			第3组			第4组		
48	71	80	42	70	77	38	73	83	49	77	84
61	74	82	48	71	81	58	74	87	58	79	93
67	75	87	62	73	89	70	75	90	73	80	94
68	79	89	67	75	92	71	79	94	74	84	97

增长的时间间隔会趋向于提高考试成绩吗？

7. 由于在低于冰点的天气发射后产生了O-环故障，挑战者号航天飞机于1986年发生了灾难性爆炸事故. 此前，火箭制造商Thiokol公司的工程师们就反对在这样冷的天气发射，因为冷天气会有发生O-环故障的危险. 他们给出了在这之前24个发射的数据如下. 问O-环事故的次数会随气温的降低增加吗？

O-环事故	气温（华氏温度）								
没有	66	67	67	67	68	68	70	70	
	72	73	75	76	76	78	79	80	81
一个	57	58	63	70	70				
两个	75								
三个	53								

详见Feynman(1988).

思考题

1. 在没有结时，证明（4）式和（5）式是ρ的等价表达式.
2. 对$n=5$，哪种秩的配对会导致

   (a) $\rho=1$？ (b) $\tau=1$？ (c) $\rho=-1$？ (d) $\tau=-1$？
3. 将思考题2中的结果推广到对于任意一般的n，证明ρ和τ事实上的确假设了指定值.
4. 假定有人建议用式子

$$R = 1 - \frac{\sum_{i=1}^{n} |R(X_i) - R(Y_i)|}{(1/4)n^2}$$

   它有时称作"Spearman's footrule".

331 (a) 在什么条件下$R=1$？ (b) 在什么条件下$R=-1$？
5. 在$n=3$和通常独立性假设的情形下，求出思考题4中的ρ、τ和R的精确分布.
6. 利用习题2中的数据，计算思考题4中定义的R. 在性质上，R能否比τ更类似于ρ吗？

5.5 非参数线性回归方法

对于我们考查二维随机变量(X,Y)的一个随机样本$(X_1,Y_1),\cdots,(X_n,Y_n)$来说，这一节和前面秩相关那一节联系紧密. 相关方法强调的是估计X和Y之间的相关度，

而回归方法用于考查 X 和 Y 之间更紧密的关系. 回归方法的一个重要目标是，我们可以从前面的观测 (X_1, Y_1) 到 (X_n, Y_n) 得到信息的基础上，只知道 X 时预测 Y 的值. 例如，X 代表大学入学成绩，Y 代表 4 年后这个学生的 GPA，对过去的学生观测可以帮助我们预测将要入学的学生在 4 年大学内的表现. 当然，Y 仍然是随机变量，所以我们不能期望仅从知道与 Y 相关的 X 的值来决定 Y，但知道 X 应该会帮助我们对 Y 做一个更好的估计.

回归方法也可用来控制试验，其中 X 可能根本就不是随机的，但试验者可能会赋以各种不同 X 的值来决定它对 Y 的影响. 例如，X 可以代表药物的被测数量，譬如降低病人血压的药物. 在一个试验中，可能会挑选几种不同水平的 X 去确定药物对 Y 的影响，Y 是病人的反应，譬如病人血压的降低.

形式上，Y 在 X 上的回归就是给定 X 的值 x 时 Y 的条件均值.

定义 5.5.1 Y 在 X 上的回归是 $E(Y|X=x)$，回归方程是 $y = E(Y|X=x)$.

如果回归方程已知，我们可以用 y 作为纵坐标，x 作为横坐标画图，将回归显示在图上. 但是回归方程很少会知道，它是基于过去的数据所做的估计. 例如，我们想预测当 $X=6$ 时的 Y 值，如果知道 $E(Y|X=6)$，我们就用它作为 Y 的预测值；否则，我们用几个 $X=6$ 或接近 6 时 Y 的观测值的样本均值或者样本中位数来作 Y 的预测值. 这时用 3.2 节和 5.7 节中所描述的方法就可以形成 $E(Y|X=6)$ 的点估计和置信 332
区间. 为了有足够的观测使得对于每一个 X 的值，可以估计 Y 在 X 上的回归，就需要很多的观测. 对于有几百个或几千个观测的大数据集，刚才提到的非参数方法会表现得很好，这是不足为奇的

当我们只有很少的观测数据，且希望估计 Y 在 X 上的回归时，这是比较困难的情形，也正是我们本节要讨论的. 知道一些 $E(Y|X=x)$ 和 x 之间的关系，而且当有很少的观测时能够用这些信息，这对我们是很有帮助的. 首先，我们将分析 $E(Y|X=x)$ 是 x 的线性函数的情形，下一节，我们考虑更一般的情形，即 $E(Y|X=x)$ 是 x 的单调（或者增或者减）函数的情形.

如果回归方程的图形是直线，则说 Y 在 X 上的回归是线性的.

定义 5.5.2 称 Y 在 X 上的回归是*线性回归*（linear regression），如果回归方程的形式是

$$E(Y|X=x) = \alpha + \beta x \tag{1}$$

对某个常数 α，称为 *y-截距*（intercept），β 称为*斜率*（slope）.

通常，常数 α 和 β 是未知的，必须用数据来估计. 如果 X 和 Y 的所有观测都用于估计 α 和 β，从而实现了数据的最大用处，同时对每个 x，可以期望一个好的 $E(Y|X=x)$ 估计. 人们广泛接受的估计 α 和 β 的一个方法被称为*最小二乘法*（least squares method）.

定义 5.5.3 在回归方程 $y=\alpha+\beta x$ 中，选择 α 和 β 的估计 a 和 b，使得最小化如下离差平方和

$$SS = \sum_{i=1}^{n} [Y_i - (a + bX_i)]^2 \tag{2}$$

其中，$(X_1, Y_1), \cdots, (X_n, Y_n)$为观测，这种估计方法称为最小二乘法.

因为真正的回归线可能和观测很接近，最小二乘方法的基本思想是回归线的估计应该是和 X 和 Y 的观测值是接近的. 因此，当同时考虑所有的点时，选择估计要使得 Y_i 和估计的回归线之间的垂直距离 D_i 很小，估计的回归线在 X_i 处的值等于 $a + bX_i$，它可能在 Y_i 之上也可能之下. 我们不能仅使得 D 的和小，因为即使估计的回
333 归线根本不接近观测，D 的和也可能是零，即距离 D 的绝对值可以很大，但正的 D 可以和负的 D 抵消，而其和为0. 为了避免这点，我们选择最小化 D 的平方和：

$$SS = \sum_{i=1}^{n} D_i^2 \tag{3}$$

其中，

$$D_i = Y_i - (a + bX_i) \tag{4}$$

这通常会提供出一条和数据吻合得很好的直线，所以是真实回归线的一个合理的估计.

▶线性回归的非参数方法

数据　数据是由来自某个二维总体的随机样本$(X_1, Y_1), \cdots, (X_n, Y_n)$组成.

假定条件

1. 样本是随机样本. 如果 X 的值是非随机量，只要 Y 是独立且条件分布相同，则这一节的方法仍然适用.

2. Y 在 X 上的回归是线性的. 这说明 X 和 Y 都是区间度量尺度.

最小二乘估计　最小二乘方法提供了真正回归线 $y = \alpha + \beta x$ 的一个估计，

$$y = a + bx \tag{5}$$

其中，a 和 b 由下式计算

$$b = \frac{n\sum_{i=1}^{n} X_iY_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} \tag{6}$$

和

$$a = \overline{Y} - b\overline{X} \tag{7}$$

334 其中，$\bar{X}$ 和 $\bar{Y}$ 是各自的样本均值.

斜率检验　为了检验关于斜率的假设，除了假定条件1和2之外，还要加上下面的假定条件.

3. "残差" $Y - E(Y|X)$ 和 X 独立.

可以采用 Spearman ρ 检验下面关于斜率的假设. 记 β_0 代表某个指定数，对每对 (X_i, Y_i)，计算 $Y_i - \beta_0 X_i = U_i$（记号）. 然后根据5.4节所述，求出数对(X_i, U_i)，$i =$

$1,\cdots,n$ 的 Spearman 秩相关系数 ρ. 表 10 给出了当 H_0 为真且没有结时 ρ 的分位数.

A.（双边检验）

$$H_0: \beta = \beta_0$$
$$H_1: \beta \neq \beta_0$$

如 5.4 节所讨论的 Spearman ρ 双边检验，如果 ρ 超过它零分布的 $1-\alpha/2$ 分位数或小于它的 $\alpha/2$ 分位数，则以水平 α 拒绝 H_0.

B.（左边检验）

$$H_0: \beta = \beta_0$$
$$H_1: \beta < \beta_0$$

如 5.4 节所讨论的 Spearman ρ 左边检验，如果 ρ 小于它零分布的 α 分位数，则以水平 α 拒绝 H_0.

C.（右边检验）

$$H_0: \beta = \beta_0$$
$$H_1: \beta > \beta_0$$

如 5.4 节所讨论的 Spearman ρ 右边检验，如果 p 大于它零分布的 $1-\alpha$ 分位数，则以水平 α 拒绝 H_0.

斜率的置信区间　这个方法也用到假定条件 1,2 和 3. 对每一对点 (X_i, Y_i) 和 (X_j, Y_j)，使得 $i<j$ 且 $X_i \neq X_j$，来计算“两点斜率”

$$S_{ij} = \frac{Y_j - Y_i}{X_j - X_i} \tag{8}$$

记 N 为所计算的斜率个数，对所得到的斜率进行排序，记

$$S^{(1)} \leqslant S^{(2)} \leqslant \cdots \leqslant S^{(N)}$$

为排序的斜率. 335

为求 β 的一个 $1-\alpha$ 置信区间，我们要找到 $w_{1-\alpha/2}$，即从表 A11 中找到 $T=N_c-N_d$ 的 $1-\alpha/2$ 分位数. 记 r 和 s 为

$$r = \tfrac{1}{2}(N - w_{1-\alpha/2}) \tag{9}$$
$$s = \tfrac{1}{2}(N + w_{1-\alpha/2}) + 1 = N + 1 - r \tag{10}$$

如果它们不是整数，对 r 下舍入整数，对 s 上舍入整数，则 β 的 $1-\alpha$ 置信区间由区间 $(S^{(r)}, S^{(s)})$ 给出，即

$$P(S^{(r)} < \beta < S^{(s)}) \geqslant 1 - \alpha \tag{11}$$

计算机辅助　*Minitab* 可以求所有成对斜率，并容易得到这个置信区间. ◀

评注

与最小二乘的概念完全不同，斜率的置信区间是基于 Kendall τ，尽管可能性不大，但 β 的最小二乘估计 b 在 β 的置信区间之外是有可能的. 这可能会发生，例如，根据其他观测来判断，当 Y 的一个值比我们期望的偏离非常大或非常小，这样的一个离群观测可以将最小二乘线“推”高（或低），使得线在损害其他观测的情况下更接近

这一点. 在这样的情形下，我们选择一个通过点（X 的样本中位数，Y 的样本中位数），斜率等于由（8）式定义的斜率 S_{ij} 的中位数的线来估计回归线更有意义. 即，我们选择估计量

$$b_1 = S_{ij}\text{的样本中位数} = S_{(\text{median})} \tag{12}$$

及

$$a_1 = Y_{0.50} - b_1 X_{0.50} \tag{13}$$

其中，$X_{0.50}$ 和 $Y_{0.50}$ 是样本中位数.

例 5.5.1

我们再次用上一节中的数据. 记 X_i 为每个 MBA 学生的 GMAT 成绩，记 Y_i 为毕业 GPA. 12 个观测 (X, Y) 是 (710, 4.0), (610, 4.0), (640, 3.9), (580, 3.8), (545,
336 3.7), (560, 3.6), (610, 3.5), (530, 3.5), (560, 3.5), (540, 3.3), (570, 3.2) 和 (560, 3.2). 这些点和最小二乘回归线

$$y = 1.4287 + 0.003714x$$

画在图 5-1 上，回归线的确定是将

$$\sum_{i=1}^{12} X_i = 7{,}015 \qquad \overline{X} = 584.58 \qquad \sum_{i=1}^{12} X_i^2 = 4{,}129{,}525$$

$$\sum_{i=1}^{12} Y_i = 43.2 \qquad \overline{Y} = 3.6 \qquad \sum_{i=1}^{12} X_i Y_i = 25{,}360.5$$

代入（6）式和（7）式，即得到了 $b = 0.003714$, $a = 1.4287$. 我们可以用回归线来作为 Y 和 X 之间关系的描述，或者更确切地说，作为给定 X, Y 的条件均值 $E(Y|X)$ 的估计. 如果一个研究生的 GMAT 成绩是 550，我们可以预测这个学生的毕业 GPA 大约是 $1.4287 + 0.003714(550) = 3.47$ 左右. 由于其他相关因素的影响，每个学生可能会有更高或更低的 GPA，诸如动机、学习习惯和竞争义务. 这个回归线估计仅提供了一个点估计.

假设有一个国家研究报告说 "GMAT 成绩增加 40 分导致 GPA 至少增加 0.4". 因为斜率是 Y 的变化量除以 X 的变化量，这个报告等价于说 GPA 对 GMAT 成绩的回归线的斜率至少是 $0.4/40 = 0.01$.

为了看我们 12 个研究生的样本是否和国家研究报告一致，我们作如下左边检验：

$$H_0: \beta \geqslant 0.01$$

337 对备择假设 $\quad H_1: \beta < 0.01$

像我们计算的那样，样本斜率小于 0.01，可以简单地认为是偶然波动的结果，并不能说明与国家调查报告不一致. 计算了 GMAT 成绩 X 和样本残差 $U = Y - (0.01)X$ 之间的 Spearman ρ.

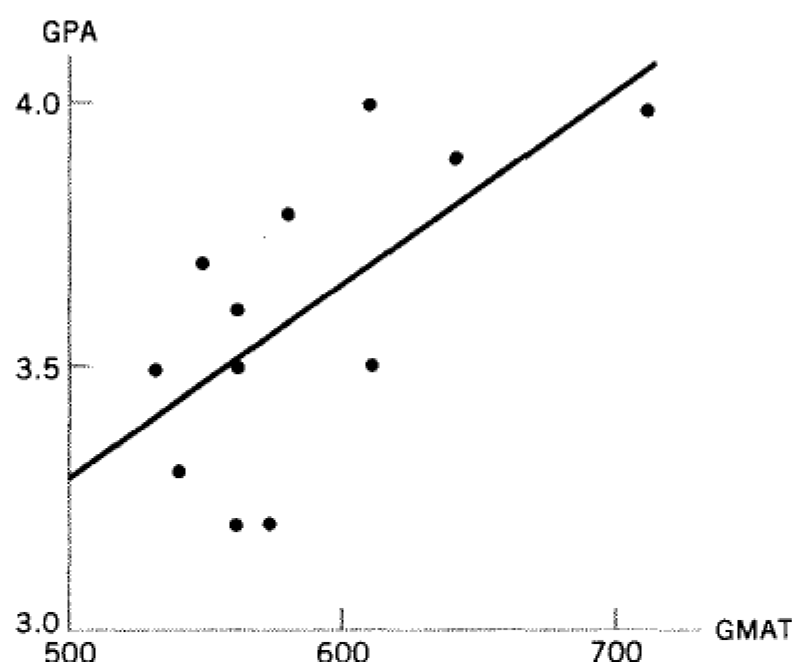


图 5-1 12 个 MBA 研究生的 GMAT 成绩对 GPA 的散点图及最小二乘回归线

	MBA 研究生 i					
	1	2	3	4	5	6
X_i	710	610	640	580	545	560
$U_i = Y - \beta_0 X_i$	−3.1	−2.1	−2.5	−2.0	−1.75	−2.0
$R(X_i)$	12	9.5	11	8	3	5
$R(U_i)$	1	7	3.5	9.5	12	9.5
	7	8	9	10	11	12
X_i	610	530	560	540	570	560
$U_i = Y - \beta_0 X_i$	−2.6	−1.	−2.1	−2.1	−2.5	−2.4
$R(X_i)$	9.5	1	5	2	7	5
$R(U_i)$	2	11	7	7	3.5	5

(5.4.4) 式用来计算 $\rho = -0.7273$，它小于表 A10 中的零分布的 0.05 分位数，所以以 $\alpha = 0.05$ 拒绝零假设. 由 (5.4.13) 式，p-值近似为

$$P(Z \leqslant -0.7273\sqrt{11}) = P(Z \leqslant -2.4121) = 0.008$$

这个 MBA 研究生的样本与国家调查结果不一致.

为了构造这个 MBA 研究生样本来自总体的真实斜率的 95% 置信区间，我们计算满足 $X_i \neq X_j$ 的所有两点斜率

$$S_{ij} = \frac{Y_j - Y_i}{X_j - X_i}$$

如图 5-2 所设计的数据表格可方便地计算 S_{ij}，共有 $N = 62$ 对 (X_i, Y_i) 和 (X_j, Y_j) 满足 $X_i \neq X_j$，如图 5-2 所示.

对 $n = 12$，从表 A11 中可找到 T 的 0.975 分位数是 28，所以由 (9) 式和 (10) 式得出，$r = 17, s = 46$. 从图 5-2 中可找到 S_{ij} 的第 17 个有序值是

$$S^{(17)} = 0.00000$$

S_{ij}的第46个有序值是

$$S^{(46)} = 0.00800$$

所以，真实斜率β的95%置信区间是从0.000到0.008.

如果由于某种原因，认为最小二乘回归线是不满足的，则我们找S_{ij}的中位数的值，它是S_{ij}的第31个有序值0.1/30和第32个有序值0.4/110的均值，即中位数是

338 0.003485，这提供了β的另一个估计，正如（12）式所描述的，那么（13）式则提供了α的估计，

$$a_1 = Y_{0.50} - b_1 X_{0.50} = 3.55 - 0.003485(565) = 1.581$$

因此，正如前面评注中所讨论的那样，回归线的另一估计为：

$$y = 1.581 + 0.003485x$$

(530, 3.5)	.00278	.00364	.00625	0	.00600	−.00750	.00333	0	−.01000	.01333	−.02000
(540, 3.3)	.00412	.00600	.01000	.00286	.01250	−.00333	.01500	.01000	−.00500	.08000	(540, 3.3)
(545, 3.7)	.00182	.00211	.00462	−.00308	.00286	−.02000	−.00667	−.01333	−.03333	(545, 3.7)	
(560, 3.2)	.00533	.00875	.01600	.00600	.03000	0	NA	NA	(560, 3.2)		
(560, 3.5)	.00333	.00500	.01000	0	.01500	−.03000	NA	(560, 3.5)			
(560, 3.6)	.00267	.00375	.00800	−.00200	.01000	−.04000	(560, 3.6)				
(570, 3.2)	.00571	.01000	.02000	.00750	.06000	(570, 3.2)					
(580, 3.8)	.00154	.00167	.00667	−.01000	(580, 3.8)						
(610, 3.5)	.00500	.01333	NA	(610, 3.5)							
(610, 4.0)	0	−.00333	(610, 4.0)								
(640, 3.9)	.00143	(640, 3.9)									
	(710, 4.0)										

图5-2　按X的增加次序排列点(X,Y)的电子数据表格以求S_{ij}的值　■

注意，在前面的例子中，斜率β的95%置信区间是从0.000到0.008，而且和拒绝零假设$\beta=0.01$的假设检验是一致的，这就是通常的情形；然而，在假设检验和置信区间之间有两个原因可能会不一致．一个原因是双边置信区间是双边假设检验的逆，而这里是单边检验．另一个原因是这个置信区间是基于Kendall τ的假设检验的逆，而这里假设检验是基于Spearman ρ的.

为了更进一步解释，我们提出的假设检验是在零假设下X和残差$U=Y-\beta_0 X$之间秩相关的非参数检验．对这个检验，我们用Spearman ρ，但是，如Theil(1950)所建议的，我们用Kendall τ作为检验统计量也是可以的．我们选择Spearman ρ是因为它更容易计算.

339 可以把我们提出的β假设检验倒过来求β的置信区间，即找出基于Spearman ρ双边检验中作为零假设所有"可接受的"β_0的值．事实上，Taylor和Conover(1988)对这个方法进行了研究，并发现会导致更多的计算，与提出的基于Kendall τ的方法

比较时，在效率上没有什么优点.

□**理论** 为推导 a 和 b，使得（2）式中的 SS 最小化，我们在括号里加减一项（$\bar{Y}-b\bar{X}$），得到

$$SS=\sum_{i=1}^{n}[(Y_i-\bar{Y})-b(X_i-\bar{X})+(\bar{Y}-b\bar{X}-a)]^2 \tag{14}$$

由代数恒等式

$$(c-d+e)^2=c^2+d^2+e^2-2cd+2ce-2de \tag{15}$$

我们并用 $c=Y_i-\bar{Y}$ 等，将（14）式展开，得

$$\begin{aligned}SS=&\sum_{i=1}^{n}(Y_i-\bar{Y})^2+b^2\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{n}(\bar{Y}-b\bar{X}-a)^2\\&-2b\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})+2(\bar{Y}-b\bar{X}-a)\sum_{i=1}^{n}(Y_i-\bar{Y})\\&-2b(\bar{Y}-b\bar{X}-a)\sum_{i=1}^{n}(X_i-\bar{X})\end{aligned} \tag{16}$$

根据 $\bar{Y}$ 和 $\bar{X}$ 的定义，$\sum(Y_i-\bar{Y})=0$ 和 $\sum(X_i-\bar{X})=0$，（16）式中的最后两个和等于 0. 当

$$a=\bar{Y}-b\bar{X} \tag{17}$$

第 3 个和达到最小（0），这给出了 a 的最小二乘解. 剩下的问题就是找使得第 2 个和与第 4 个和达到最小化的 b 值，即最小化

$$b^2S_x-2bS_{xy} \tag{18}$$ 340

其中

$$S_x=\sum_{i=1}^{n}(X_i-\bar{X})^2 \tag{19}$$

及

$$S_{xy}=\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) \tag{20}$$

在（18）式中，加减一项 S_{xy}^2/S_x，则第 2 个和与第 4 个和是

$$S_x\left[b^2-2b\frac{S_{xy}}{S_x}+\left(\frac{S_{xy}}{S_x}\right)^2\right]-\frac{S_{xy}{}^2}{S_x}=S_x\left(b-\frac{S_{xy}}{S_x}\right)^2-\frac{S_{xy}{}^2}{S_x}$$

当

$$b=\frac{S_{xy}}{S_x} \tag{21}$$

它显然是最小值，且与（6）式一致. 注意，这时第 2 个和与第 4 个和变成 $-S_{xy}^2/S_x$，所以平方和的最小值是：

$$SS_{\min}=\sum_{i=1}^{n}(Y_i-\bar{Y})^2-\frac{S_{xy}{}^2}{S_x}=(1-r^2)\sum_{i=1}^{n}(Y_i-\bar{Y})^2 \tag{22}$$

这里r是由（5.4.1）式给出的 Pearson 乘积矩相关系数．注意，关于(X,Y)的分布我们没有作任何假设，所以最小二乘方法是分布自由的．事实上，假定条件 1 和 2 的唯一目的是保证我们所估计的回归线存在.

在假定条件 3 下，残差

$$Y_i - E(Y_i|X_i) = Y_i - (\alpha + \beta X_i) \tag{23}$$

和X_i是独立的，所以 5.4 节中关于 Spearman ρ 的假定条件成立．注意，$(Y_i-\alpha-\beta X_i)$，$i=1$ 到 n 的秩和 $U_i=(Y_i-\beta X_i)$，$i=1$ 到 n 的秩是相同的，所以我们可以在不知道α的情况下，检验H_0：$\beta=\beta_0$，就像 Spearman ρ 仅是用秩来计算的 Pearson r 一
341 样，这个检验类似于在(X_i,U_i)上计算r的秩检验，这是通常检验同一零假设的参数方法，即是在假设(X,Y)有二维正态分布下的参数方法．在与前面相同条件以及X的观测是等间隔的条件下，根据 Stuart(1954,1956)，这个方法的渐近相对效率(A. R. E.)是$(3/\pi)^{1/3}=0.98$；对其他分布，A. R. E. 经常大于或等于 0.95(Lehmann,1975).

为了弄清斜率S_{ij}和 Kendall τ之间的关系，注意，对假设的斜率β_0，我们有

$$\begin{aligned} S_{ij} &= \frac{Y_i - Y_j}{X_i - X_j} = \frac{U_i + \beta_0 X_i - U_j - \beta_0 X_j}{X_i - X_j} \\ &= \beta_0 + \frac{U_i - U_j}{X_i - X_j} \end{aligned} \tag{24}$$

其中，$U_i=Y_i-\beta_0X_i-\alpha$ 是 Y_i 与假设回归线 $y=\alpha+\beta_0x$ 之间的残差．依照(X_i,U_i)和(X_j,U_j)是协调或者是不协调（在 5.4 节对 Kendall τ 的讨论中描述过）来确定斜率S_{ij}大于β_0或小于β_0．如果我们用S_{ij}小于β_0的个数来作检验统计量，决定是否接受H_0：$\beta=\beta_0$，那么只要不协调对的个数N_d不是太小或太大，我们就接受β_0．因为N_d和协调对的个数N_c有下式关系：

$$N_c + N_d = N \tag{25}$$

其中，N是总对数，而且因为如果我们有真实斜率和独立性的假定条件 3，则表 A11 可给出N_c-N_d的分位数．那么如果N_c-N_d大于表 A11 中的$w_{1-\alpha/2}$，我们就说N_d太小．这也等价于说N_d小于$r=(N-w_{1-\alpha/2})/2$，换句话说，如果β_0至少大于r个S_{ij}（或者$\beta_0>S^{(r)}$），则β_0是可接受的．同样的讨论可给出β_0的上界，从而得到置信区间．这个方法源于 Theil(1950)，Sen(1968a)把它修改为处理有结时的情况．□

对非参数检验适用于几种回归线的情形，可参见 Sen(1972)，Adichie(1974，1975)和 Pothoff(1974)．Jureckova(1971，1977)，Huber(1973)以及 Hettmansperger 和 McKean(1977)给出了其他的估计回归系数的方法，Kalbfleish(1974)讨论了非线性模型中的秩方法．有关非参数回归的更进一步讨论见 Puri(1985)，Jaeckel(1972)，Hollander 和 Wolfe(1973)，Behnen(1976)以及 Stone(1977).

习题

342 1.　一名汽车司机记录了她的行驶里程数和她每次加油的加仑数.

英里	加仑	英里	加仑
142	11.1	157	12.5
116	5.7	255	17.9
194	14.2	159	8.8
250	15.8	43	3.4
88	7.5	208	15.2

(a) 用加仑作 x 轴，作图显示这些点.

(b) 用最小二乘方法估计 a 和 b.

(c) 在 (a) 的图中画出最小二乘回归线.

(d) 假设 EPA 以每加仑 18 英里估计这辆车的英里数. 检验零假设：这个数适用于这辆特定的车和司机（用斜率检验）.

(e) 求出这辆车和司机的英里数的 95% 置信区间.

2. 美国学院和大学的学生和教师人数（1973 年春季）的一个随机样本如下：

名字	学生	教师
American International	2546	129
Bethany Nazarene	1355	75
Carlow	1019	87
David Lipscomb	1858	99
Florida International University	4500	300
Heidelberg	1141	109
Lake Erie	784	77
Mary Hardin Baylor	1063	64
Mt. Angel	267	40
Newberry	753	61
Pacific Lutheran University	3164	190
St. Ambrose	1189	90
Smith	2755	240
Texas Women's University	5602	300
West Liberty State	2697	170
Wofford	988	73

(a) 用教师作 x 轴，作图显示这些点.

(b) 用最小二乘方法估计回归线.

(c) 在 (a) 的图中画出最小二乘回归线.

(d) 检验假设：每增加一名教师，就伴随平均增加 15 个学生.

(e) 求出斜率的置信区间.

3. 考查研究生院申请者的一个随机样本. 检验零假设：GRE 语文成绩（Y）和 GRE 数学成绩（X）之间的线性回归有斜率 1.0，备择假设是斜率小于 1.0. 343

学生	数学	语文	学生	数学	语文
1	650	540	9	460	510
2	720	580	10	520	500
3	580	500	11	740	680
4	670	570	12	450	600
5	600	630	13	530	550
6	510	630	14	570	500
7	480	520	15	680	510
8	610	610	16	740	570

4. 一个公司的共有股票的变化率定义为它的表现（Y）与标准普尔500指数的表现（500家股票的一个指数）相比较的斜率. 对最后8个季度测量了它们的表现. 试求这个斜率的95%置信区间.

季度	公司(X)	标准普尔500指数	季度	公司(X)	标准普尔500指数
1	+4.5%	+2.6%	5	−4.6%	−2.5%
2	+5.1%	+2.7%	6	−0.6%	+0.1%
3	+8.0%	+3.1%	7	+10.3%	+4.9%
4	+2.2%	+0.8%	8	+2.2%	+1.0%

5.6 单调回归方法

5.5节中提出了线性回归的非参数方法，这些可以用于如例5.5.1的情形，线性回归的假设看起来是合理的，但在其他情形，回归函数假设是一条直线可能并不合理，而假设随着 X 的增加，$E(Y|X)$也增加（至少它不下降）可能是合理的. 在这种情况下，我们称回归是*单调递增的*（monotonically increasing）. 如果随着 X 的增加 $E(Y|X)$减少，则称回归是*单调递减的*（monotonically decreasing）. 这两种情形都可以用下面的方法.

▶单调回归的非参数方法

数据 数据是由来自某个二维分布的随机样本（$(X_1,Y_1),\cdots,(X_n,Y_n)$）组成.

假定条件

1. 样本是随机样本.
2. Y 在 X 上的回归是单调的.

344 **$E(Y|X)$ 的点估计** 给定值 $X=x_0$，估计 Y 在 X 上的回归：

1. 获得 X 的秩 $R(X_i)$和 Y 的秩 $R(Y_i)$，有结时用平均秩.
2. 基于秩求最小二乘回归线.

$$y = a_2 + b_2x \tag{1}$$

其中

$$b_2 = \frac{\sum_{i=1}^{n} R(X_i)R(Y_i) - n(n+1)^2/4}{\sum_{i=1}^{n} [R(X_i)]^2 - n(n+1)^2/4} \tag{2}$$

及

$$a_2 = (1-b_2)(n+1)/2 \tag{3}$$

3. 获得 x_0 的秩 $R(x_0)$如下：
   (a) 如果 x_0 等于某个观测 X_i，令 $R(x_0)$等于 X_i 的秩.
   (b) 如果 x_0 落入两个邻近的值 X_i 和 Y_j，这里，$X_i<x_0<X_j$，对它们各自的秩做内插，得到 $R(x_0)$.

$$R(x_0) = R(X_i) + \frac{x_0 - X_i}{X_j - X_i}[R(X_j) - R(X_i)] \tag{4}$$

这个“秩”不必是个整数

(c) 如果 x_0 小于 X 的最小观测或者大于 X 的最大观测，不要外推，因为 Y 在 X 上的回归的信息只在观测 X 的范围内有效.

4. 将 $R(x_0)$ 作为 x 代入（1）式，得到相应值 $E(Y|X=x_0)$ 估计的秩 $R(y_0)$.

$$R(y_0) = a_2 + b_2 R(x_0) \tag{5}$$

5. 通过用 Y 的观测，将 $R(y_0)$ 转换为 $\hat{E}(Y|X=x_0)$，即 $E(Y|X=x_0)$ 的估计，方法如下：

(a) 如果 $R(y_0)$ 等于某个观测 Y_i 的秩，令这个估计 $\hat{E}(Y|X=x_0)$ 等于这个观测 Y_i.

(b) 如果 $R(y_0)$ 落入两个邻近的值 Y_i 和 Y_j 之间（$Y_i < Y_j$），使得 $R(Y_i) < R(y_0) < R(Y_j)$，在 Y_i 和 Y_j 之间做内插：

$$\hat{E}(Y|X=x_0) = Y_i + \frac{R(y_0) - R(Y_i)}{R(Y_j) - R(Y_i)}(Y_j - Y_i) \tag{6}$$ 345

(c) 如果 $R(y_0)$ 大于 Y 的最大观测的秩，令 $\hat{E}(Y|X=x_0)$ 等于 Y 的最大的观测值. 如果 $R(y_0)$ 小于 Y 的最小观测的秩，令 $\hat{E}(Y|X=x_0)$ 等于 Y 的最小观测值.

Y 在 X 上的回归估计 如刚描述的方法，利用所有观测值来做回归点估计，为得到整条回归曲线，我们用下面的方法.

1. 对从 $X^{(1)}$ 到 $X^{(n)}$ 的每个 X_i，用前面描述的方法估计 $E(Y|X)$.

2. 对 Y 的每个秩，$R(Y_i)$，从（1）式，求出 X_i 的估计秩，$\hat{R}(X_i)$.

$$\hat{R}(X_i) = [R(Y_i) - a_2]/b_2, \qquad i = 1, 2, \cdots, n \tag{7}$$

3. 用前面第 5 步的方法，将每个 $\hat{R}(X_i)$ 转换为估计 $\hat{X}_i$，确切地说：

(a) 如果 $\hat{R}(X_i)$ 等于某个观测 X_j 的秩，令 $\hat{X}_i$ 等于这个观测值.

(b) 如果 $\hat{R}(X_i)$ 落入两个邻近值 X_j 和 X_k 的秩之间（$X_j < X_k$），则用内插，

$$\hat{X}_i = X_j + \frac{\hat{R}(X_i) - R(X_j)}{R(X_k) - R(X_j)}(X_k - X_j) \tag{8}$$

(c) 如果 $\hat{R}(X_i)$ 小于 X 的最小观测的秩或大于 X 的最大观测的秩，则没有估计 $\hat{X}_i$.

4. 在图纸上画出在第 1 步到第 3 步中求得的每一个点. 即，画每个（$X_i, \hat{Y}_i$）和每个（$\hat{X}_i, Y_i$）. 所有这些点应该是单调的，如果 $b_2 > 0$，则递增，如果 $b_2 < 0$，则递减.

5. 将第 4 步中相邻的两点用直线连起来，这一系列连起来的线段就构成 Y 在 X 上的回归曲线的估计. ◀

例 5.6.1

研究 17 桶新鲜葡萄汁，看葡萄汁要多久才能变成葡萄酒，并把它作为给汁里加多少糖的函数. 范围从 0 到 10 磅不同量的糖加入到那些桶中，每天都要检查那些桶，看是否转变成葡萄酒. 30 天后，试验终止，有 3 桶仍未发酵. 需要做 Y（直到发酵的天数）在 X（糖的磅数）上的回归曲线的估计

对于观测（X_i, Y_i），由前面的第 1，2 和 3 步可计算它们的秩 $R(X_i)$ 和 $R(Y_i)$，

346 值 $\hat{R}(Y_i), \hat{Y}_i = \hat{E}(Y|X_i), \hat{R}(X_i)$ 以及 $\hat{X}_i$.

X_i	Y_i	$R(X_i)$	$R(Y_i)$	$\hat{R}(Y_i)$	$\hat{Y}_i$	$\hat{R}(X_i)$	$\hat{X}_i$
0	>30	1	16	16.47	>30	1.50	.25
.5	>30	2	16	15.54	29.54	1.50	.25
1.0	>30	3	16	14.60	28.60	1.50	.25
1.8	28	4	14	13.67	26.67	3.64	1.52
2.2	24	5	13	12.74	22.67	4.71	2.09
2.7	19	6	12	11.80	18.60	5.78	2.59
4.0	17	7.5	11	10.40	15.00	6.85	3.44
4.0	9	7.5	8	10.40	15.00	10.06	5.63
4.9	12	9	9.5	9.00	11.00	8.46	4.58
5.6	12	10	9.5	8.07	9.13	8.46	4.58
6.0	6	11	5	7.13	8.13	13.28	7.50
6.5	8	12	7	6.20	7.20	11.13	6.07
7.3	4	13	1.5	5.27	6.26	17.03	缺失
8.0	5	14	3	4.33	5.67	15.42	9.01
8.8	6	15	5	3.40	5.20	13.28	7.50
9.3	4	16	1.5	2.46	4.64	17.03	缺失
9.8	6	17	5	1.53	4.02	13.28	7.50

图 5-3 求单调回归曲线估计的计算

图 5-3 中给出了这些值. 在得到 $\hat{R}(Y_i), \hat{Y}_i, \hat{R}(X_i)$ 和 $\hat{X}_i$ 之前，关于秩的最小二乘系数可从（2）式和（3）式计算得到，并代入（1）式得到关于秩的最小二乘回归线.

$$y = 17.4037 - 0.9337x \tag{9}$$

在图 5-4 中画出了观测值，同时也画出了由连接相邻两点（$X_i, \hat{Y}_i$）和（$\hat{X}_i, Y_i$）的线段所组成的回归曲线. 在图 5-4 中，由横坐标 x_0 所对应曲线上纵坐标的值就容易得到估计 $\hat{E}(Y|X = x_0)$. 注意，在秩回归中用到了"删失的"观测">30"，但是数据的回归曲线那部分不可能用普通的线性回归来画

有趣的是，对一个显然是非线性回归曲线的观测数据集，它是如何转换到看上去回归曲线为直线的秩回归上去的. 图 5-5 中画出了这些秩及其方程（9）. ■

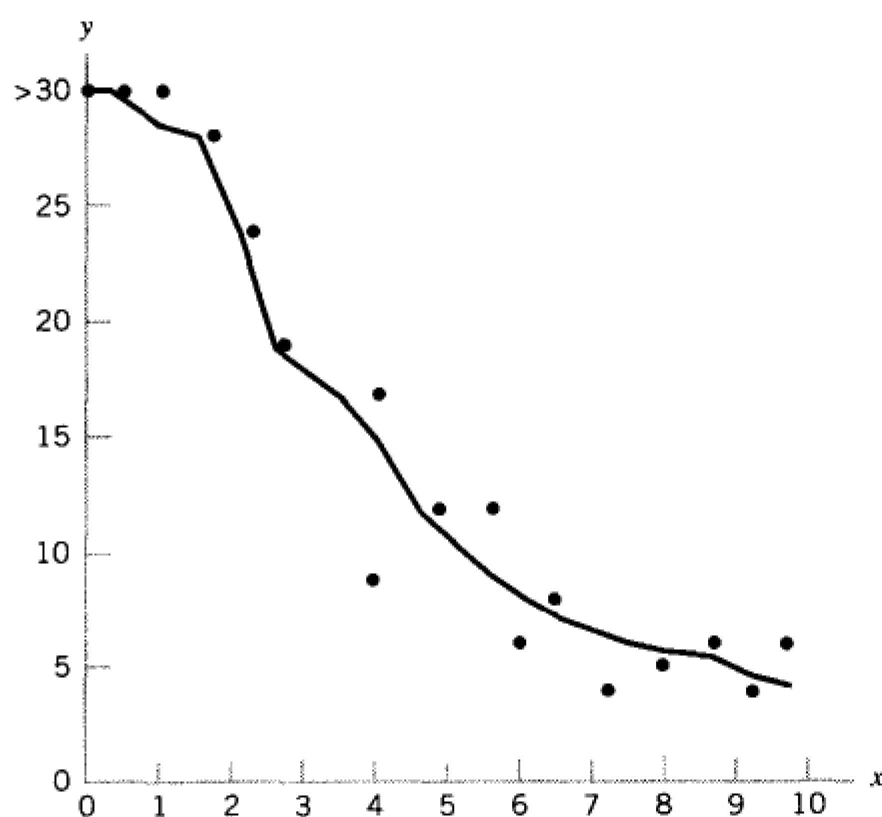


图 5-4 直到发酵的天数（y）对糖的磅数（x），以及估计的单调回归曲线

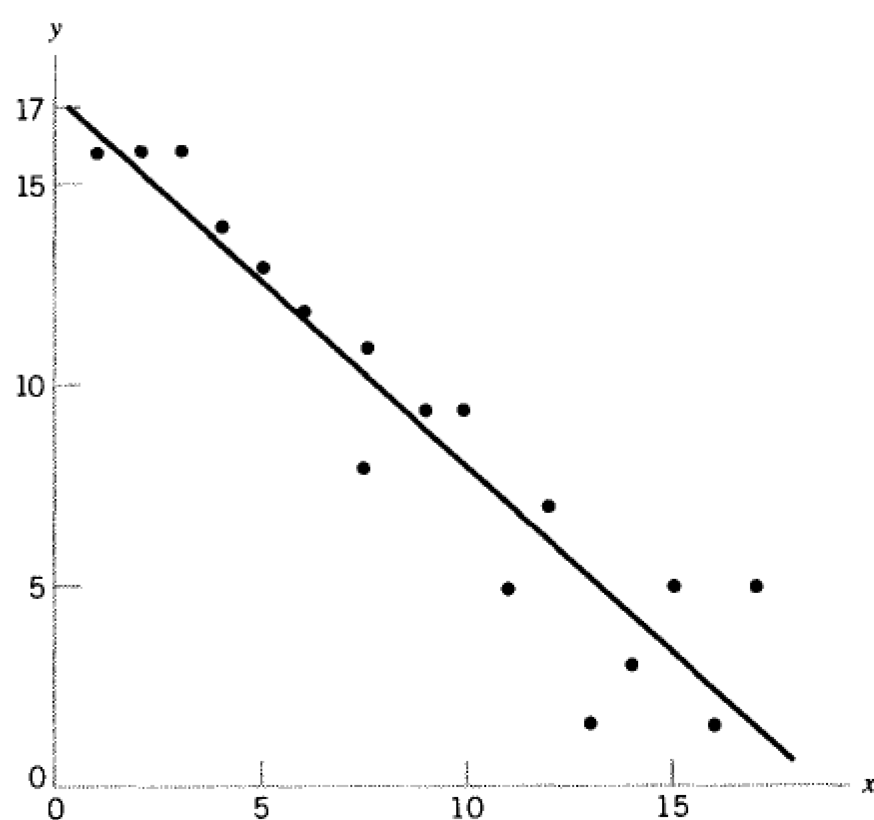


图 5-5 $R(Y_i)$对$R(X_i)$以及最小二乘回归线

□**理论** 单调回归的方法是基于如果两个变量有单调关系，那么它们的秩将会有线性关系这一基础的. 在单调回归线附近观测点的散布对应着在秩线性回归线周围秩的散布. 秩作为转换变量，所找的变换将单调回归函数转换为线性回归函数. $E(Y|X)$的区间估计可以用2.2节中描述的自助法得到. □ 347~348

Cryer，Robertson，Wright 和 Casady (1972)，Casady 和 Cryer (1976)，Hogg (1975)以及 Iman 和 Conover(1979)比较和解释过其他处理单调回归的方法．对这个方法更完整地阐述，参见 Iman 和 Conover(1979)．

习题

1. "剂量-反应"曲线广泛应用于生物学的研究和制药工业中，如下述例子．假设将某种药品（X，以毫升度量）用于天竺鼠，看是否有某种特殊反应（癌症，糖尿病，等等）发生．用5只天竺鼠做试验，每只给几个剂量水平的药物，动物显示反应的百分比作为 Y 变量，记录如下．

X(剂量)	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Y(反应百分比)	0	0	20	0	40	60	40	80	100	100

   (a) 画出散点图．期望的反应值看起来是剂量的线性函数吗？是单调函数吗？
   (b) 在 $X=3.0$ 毫升处，估计 $E(Y|X)$．
   (c) 在 $X=3.3$ 毫升处，估计 $E(Y|X)$．
   (d) 估计 Y 在 X 上的回归曲线，将估计出的回归曲线画在（a）的那张图上．

2. 10个公司公布了它们去年相比于前年的广告费用的增长百分比（X），和它们的销售额增长百分比（Y）．

	公司									
	1	2	3	4	5	6	7	8	9	10
X(广告)	4	62	31	−11	47	88	16	−1	74	21
Y(销售)	10	33	39	−14	37	39	18	−8	45	33

   (a) 画出散点图．期望的销售额增长百分比的值看起来是广告费用增长百分比的线性函数吗？是单调函数吗？
   (b) 对25%的广告费用增长，估计销售额的期望增长百分比．
   (c) 估计 Y 在 X 上的回归曲线．将估计的回归曲线画在（a）的那张图上．

3. 在确定地雷爆炸概率的试验中，给定某种刺激强度，测试17个地雷，给每个地雷以不同强度的冲撞刺激，看地雷是否爆炸．结果有8个地雷爆炸，9个没有爆炸．各自的冲撞刺激强度如下给出．
349

爆炸的	10.7, 13.9, 15.8, 17.0, 18.1, 19.9, 20.7, 21.6
没有爆炸的	4.0, 4.4, 4.7, 5.1, 9.3, 11.2, 13.7, 15.0, 19.7

   给定冲撞刺激强度为20，用单调回归估计地雷爆炸的概率（提示：如果地雷没有爆炸，则令 $Y=0$，如果地雷爆炸，则令 $Y=1$）．

思考题

1. 证明 $E(Y|X)$ 的估计不可能小于 Y 的最小观测值或大于 Y 的最大观测值．对于习题1和2所描述的情形，试讨论这个性质的优点和缺点．
2. 对习题1中的数据求出最小二乘回归直线．用这个回归直线估计给定 $X=0.5$毫升时 Y 的均值．对你来说，这个估计合理吗？

5.7 一样本或配对情形

这一节的秩检验用于处理单个随机样本和配对随机样本，当考虑差时，配对随机样本就变为单个样本. 事实上，一个配对（X_i, Y_i）是二维随机变量的单个观测. 3.4 节的符号检验通过将每对变为加，减，或一个结，并将二项检验应用到所得的单个样本上来分析配对数据. 本节的检验也用下面的差将配对（X_i，Y_i）变为单个观测.

$$D_i = Y_i - X_i \qquad \text{对 } i = 1, 2, \cdots, n \tag{1}$$

然后将 D_i 作为单个观测进行分析. 符号检验仅注重 D_i 是正的，负的，或零，而本节的检验注重正的 D_i 相对于负的 D_i 的大小. 本节的模型类似于符号检验中所用的模型，而且假设也类似于符号检验中的假设. 符号检验和这个检验之间重要的不同是关于差分布的对称性（symmetry）假定. 在介绍这个检验之前，我们应该清楚形容词"对称的"应用到分布时的意义，并讨论对称性对度量尺度的影响.

如果分布是离散的，则对称性容易定义. 如果离散型概率函数图的左半边是右半边的镜像，则称离散型分布是对称的. 例如，如果 $p = 1/2$，则二项分布是对称的（见图5-6），而且离散的均匀分布总是对称的（见图5-7）. 图中的虚线代表分布关于此线对称. 350

对不是离散的分布，我们不能画一幅概率函数图. 因此需要一个抽象的对称性的定义，见下面的定义 5.7.1.

定义 5.7.1 对于某个 c，称随机变量 X 的分布是关于线 $x = c$ 对称的，如果对每个可能的 x 的值，$X \leqslant c - x$ 的概率等于 $X \geqslant c + x$ 的概率.

在图 5-6 中，$c = 2$，且对所有的实数 x，这个对称性定义容易验证. 在图 5-7 中，$c = 3.5$. 即使我们不知道一个随机变量的精确分布，但我们经常能说，"假设分布对称是合理的."这样的一个假设并没有正态分布假设那么强；因为所有的正态分布都是对称的，但并不是所有的对称分布都是正态的.

如果一个分布是对称的，则均值（如果它存在）与中位数一致，因为两个都处在分布的中间，在对称线处. 给模型加对称性假设的一个结果就是任何关于中位数的推断对均值的推断也有效.

给模型加对称性假设的第二个结果就是要求的度量尺度从有序的变为区间的. 对于次序度量尺度，随机变量的两个观测只需要基于谁大谁小来区别，不必要知道哪一个离中位数最远，例如，当两个观测在中位数的两边. 如果对称性的假设是有意义的，离中位数的距离就是有意义的度量，从而，两个观测之间的距离就是有意义的度量. 所以，度量尺度不仅是有序的，而且它是区间的.

Wilcoxon(1945)提出了一个检验，它设计为检验一个特殊的样本是否来自一个指定均值或中位数的总体，它也可以用于观测是配对的情形，如通过每个项目"之前"和"之后"的观测来看对配中第二个随机变量是否和第一个随机变量有

相同的均值．注意，在对称分布中，均值等于中位数，所以这两个术语可以互相交换使用．

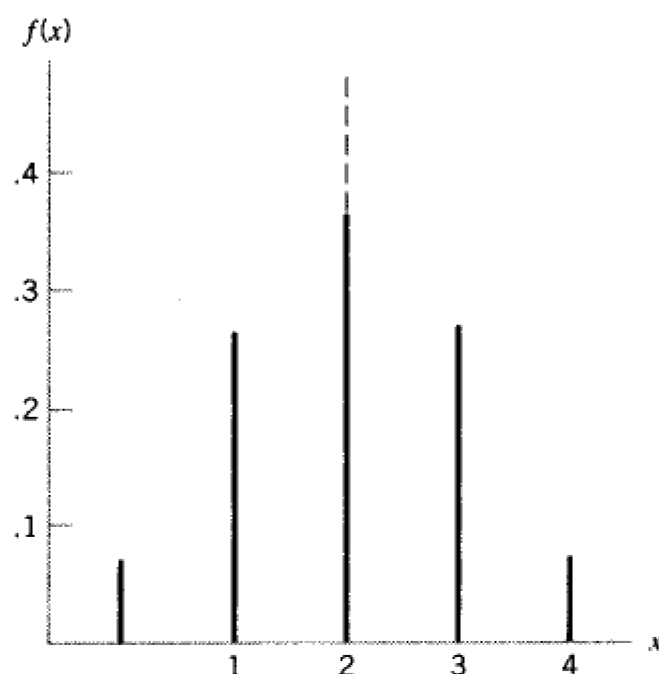


图5-6　二项分布的对称性

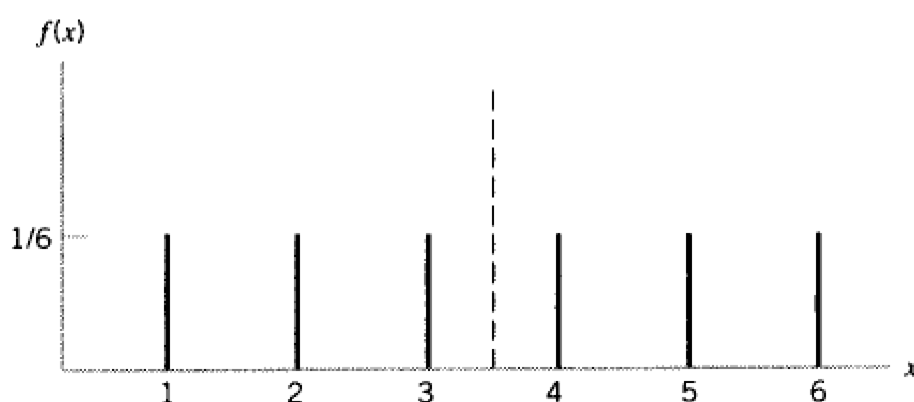


图5-7　离散均匀分布的对称性

▶Wilcoxon 符号秩检验

数据　数据由 n' 个独立观测值 $(x_1,y_1),(x_2,y_2),\cdots,(x_{n'},y_{n'})$ 组成，其中 (x_i,y_i) 是二维随机变量 (X_i,Y_i) 的观测值，$i=1,2,\cdots,n'$．求出 n' 个差 $D_i=Y_i-X_i$（在一样本问题中，D 是样本的观测，如例5.7.2中的解释）．然后对 n' 个数对 (X_i,Y_i) 的每一数对计算绝对差（与符号无关）

$$|D_i|=|Y_i-X_i| \qquad i=1,2,\cdots,n' \tag{2}$$

省略对所有差为零（即 $X_i=Y_i$，或者 $D_i=0$）的数对的进一步考虑，令数对的个数仍由 n 来记，$n\leqslant n'$．根据绝对差的大小，把从1到 n 的秩如下赋给这 n 个数对．秩1赋给绝对差 $|D_i|$ 最小的数对 (X_i,Y_i)；秩2赋给绝对差第二小的数对；等等，秩 n 赋给绝对差最大的数对．

351 如果几个数对的绝对差互相相等，则给每个数对赋以本该赋给它们秩的平均秩．
352 [即如果秩3,4,5和6属于4个数对，但是我们不知道将哪个值赋给哪个数对，因为

所有 4 个数对的绝对差都相互相等，那么给这 4 个数对都赋以平均秩$\frac{1}{4}(3+4+5+6)=4.5$].

假定条件

1. 每个 D_i 的分布都是对称的.
2. D_i 相互独立.
3. 所有 D_i 都有相同的均值.
4. D_i 的度量尺度至少是区间的.

检验统计量 对每个数对 (X_i, Y_i)，定义符号秩（记为 R_i）如下.

$$R_i = \text{赋给}\ (X_i, Y_i)\ \text{的秩, 如果}\ D_i = Y_i - X_i\ \text{是正的(即},\ Y_i > X_i)$$
$$R_i = \text{赋给}\ (X_i, Y_i)\ \text{的秩为负数, 如果}\ D_i = Y_i - X_i\ \text{是负的(即},\ Y_i < X_i)$$

检验统计量是正符号秩的和

$$T^+ = \sum_{D_i > 0} R_i \tag{3}$$

零分布 在 D_i 有均值 0 的零假设下，对没有结，且 $n \leqslant 50$ 时，表 A21 给出了 T^+ 精确分布的下侧分位数. 上侧分位数可从下面关系式中获得

$$w_p = \frac{n(n+1)}{2} - w_{1-p} \tag{4}$$

如果有很多结时，或者如果 $n > 50$，则应当用正态逼近. 正态逼近要用所有带有正或负号的符号秩的和，以及统计量

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{\sum_{i=1}^{n} R_i^2}} \tag{5}$$

在没有结的情况下，根据引理 1.4.2，(5) 式可简化为

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{n(n+1)(2n+1)/6}} \tag{6}$$ 353

T 的零分布近似于标准正态（见表 A1）.

假设

A.（双边检验）

$$H_0: E(D) = 0 \qquad (\text{即},\ E(Y_i) = E(X_i))$$
$$H_1: E(D) \neq 0$$

如果对 (X_i, Y_i) 有相同的二维分布，那么，H_1 可以写为：$E(X) \neq E(Y)$. 如果 T^+ (或 T) 小于它零分布的 $\alpha/2$ 分位数或大于它零分布的 $1-\alpha/2$ 分位数，则以水平 α 拒绝 H_0，T^+ 的分位数可在表 A12 中获得. T 的近似分位数可从表 A1 中获得.

双边检验的 p-值是 2 倍的单边 p-值中较小者，或者从正态分布近似

$$\text{左边}p\text{-值}=P\left(Z\leqslant\frac{\sum_{i=1}^{n}R_i+1}{\sqrt{\sum_{i=1}^{n}R_i^2}}\right) \tag{7}$$

或者

$$\text{右边}p\text{-值}=P\left(Z\geqslant\frac{\sum_{i=1}^{n}R_i-1}{\sqrt{\sum_{i=1}^{n}R_i^2}}\right) \tag{8}$$

B.（左边检验）

$$H_0: E(D)\geqslant 0 \qquad (\text{即}, E(Y_i)\geqslant E(X_i))$$
$$H_1: E(D)<0$$

如果数对（X_i, Y_i）有相同的二维分布，那么，H_1 可以写为 $E(Y) < E(X)$. 如果 T^+（或 T）小于它零分布的 α 分位数（对于 T^+，分位数可从表 A12 中获得；对于 T，分位数可从表 A1 中获得），则以水平 α 拒绝 H_0. 近似左边 p-值的由（7）式给出.

C.（右边检验）

$$H_0: E(D)\leqslant 0 \qquad (\text{即}, E(Y_i)\leqslant E(X_i))$$
354 $$H_1: E(D)>0$$

如果数对（X_i, Y_i）有相同的二维分布，那么，H_1 可以写为：$E(Y) > E(X)$. 如果 T^+（或 T）大于它零分布的 α 分位数（对于 T^+，分位数可从表 A12 中获得，对于 T，分位数可从表 A1 获得），则以水平 α 拒绝 H_0. 近似右边 p-值由（8）式给出.

计算机辅助 *Minitab*, *S-Plus*, *SAS* 和 *StatXact* 含有作 Wilcoxon 符号秩检验的程序. ◀

例 5.7.1

给 12 组双胞胎做心理检验，以测量每个人的进取心. 我们感兴趣的是对双胞胎进行比较，看第一个出生的是否倾向于比另外一个更有进取心. 结果如下，高分显示更多的进取心.

双胞胎组

	1	2	3	4	5	6	7	8	9	10	11	12
第一个出生 X_i	86	71	77	68	91	72	77	91	70	71	88	87
第二个出生 Y_i	88	77	76	64	96	72	65	90	65	80	81	72
差 D_i	+2	+6	−1	−4	+5	0	−12	−1	−5	+9	−7	−15
$\lvert D_i\rvert$ **的秩**	3	7	1.5	4	5.5	—	10	1.5	5.5	9	8	11
R_i	3	7	−1.5	−4	5.5	—	−10	−1.5	−5.5	9	−8	−11

假设是

H_0: 双胞胎中第一个出生的并不比第二个更具有进取心 $(E(X_i) \leqslant E(Y_i))$

H_1: 双胞胎中第一个出生的比第二个出生的更具有进取心 $(E(X_i) > E(Y_i))$

这些对应于假设 B. 我们假设测试分数是个人进取心的精确度量. 由于有很多结，检验统计量是

$$T = \frac{\sum R_i}{\sqrt{\sum R_i^2}} = \frac{-17}{\sqrt{505}} = -0.7565 \tag{9}$$

水平 $\alpha = 0.05$ 的判别区域对应于着 T 值小于 -1.6449（由表 A1 获得），因此，接受 H_0. 由（7）式，p-值是 0.238.

如果用 T^+ 和表 A12，我们会得到 $T^+ = 24.5$，临界域对应着 T^+ 的值小于 14. 所以得到同样的结论，类似的 p-值由对表 A12 中的 $w_{0.20}$ 和 $w_{0.30}$ 插值得到. ■

Wilcoxon 符号秩检验等同于中位数检验，其中数据由容量为 n' 的单个随机样本 $Y_1, \cdots, Y_{n'}$ 组成. 令 Y 是与 Y_i 同分布的随机变量，m 是指定的常数. 对应于前面假设 355
集 A，B 和 C 的假设如下.

(a)（双边检验）

H_0: Y 的中位数等于 m

H_1: Y 的中位数不等于 m

(b)（左边检验）

H_0: Y 的中位数 $\geqslant m$

H_1: Y 的中位数 $< m$

(c)（右边检验）

H_0: Y 的中位数 $\leqslant m$

H_1: Y 的中位数 $> m$

由于 Y 的分布对称性假设，"均值" 可以代替这些假设的 "中位数".

这就形成了数对 $(m, Y_1), (m, Y_2), \cdots, (m, Y_{n'})$，完全按照 Wilcoxon 符号秩检验描述的来处理这些数对，而 Wilcoxon 检验方法中其余部分保持不变. 下面的例子来解释这个方法.

例 5.7.2

为了检验假设 Y 的均值，$E(Y)$ 不会大于 30（假设 C），得到随机变量 Y 的 30 个观测.

H_0: $E(Y) \leqslant 30$

H_1: $E(Y) > 30$

观测值，差 $Y_i - m$，以及数对的秩列表如下（为了方便，随机样本先排序）.

Y_i	$D_i=Y_i-30$	$\|D_i\|$的秩	Y_i	$D_i=Y_i-30$	$\|D_i\|$的秩
23.8	−6.2	17	35.9	+5.9	15
26.0	−4.0	11	36.1	+6.1	16
26.9	−3.1	8	36.4	+6.4	18
27.4	−2.6	6	36.6	+6.6	19
28.0	−2.0	5	37.2	+7.2	20
30.3	+0.3	1	37.3	+7.3	21
30.7	+0.7	2	37.9	+7.9	22
31.2	+1.2	3	38.2	+8.2	23
31.3	+1.3	4	39.6	+9.6	24
32.8	+2.8	7	40.6	+10.6	25
33.2	+3.2	9	41.1	+11.1	26
33.9	+3.9	10	42.3	+12.3	27
34.3	+4.3	12	42.8	+12.8	28
34.9	+4.9	13	44.0	+14.0	29
35.0	+5.0	14	45.8	+15.8	30

356

从表A12中可得0.05分位数是152，所以0.95分位数是$465-152=313$. 因此，尺度≤0.05的临界域对应着检验统计量大于313.

用（3）式定义检验统计量，在这种情况下，T^+等于D_i为正的那些秩的和.

$$T^+=418 \tag{10}$$

由于T^+值较大，所以拒绝H_0. 我们得出结论是，Y的均值大于30.

由（8）式，近似的p-值为：

$$P\left(Z\geqslant\frac{\sum R_i-1}{\sqrt{n(n+1)(2n+1)/6}}\right)=P\left(Z\geqslant\frac{371-1}{\sqrt{9455}}\right)=P(Z\geqslant 3.8051)$$

由表A1显示p-值小于0.0001. ■

□**理论** 这个模型说明所有的差D_i有一个共同的中位数$d_{0.50}$，当H_0为真时，$d_{0.50}$等于0. 由对称性的定义，每个D_i为负的概率等于它为正的概率，对于连续分布，或者对于D_i为没有零值的离散分布，这个概率等于0.5.（若没有对称性，正的差就有可能倾向于比负的差大或者小）.

考虑这些的目的就是寻找H_0为真时检验统计量T^+的分布. 首先，我们将考虑双边检验的零假设，所得到的分布也可以相当好地应用到单边检验中.

考虑n个从1到n编号的筹码，如果数据没有结，其对应于数据的n个秩. 假设每个筹码的编号都写到它的一面（正面），而它的编号的负数写到另一面（反面）（像6和−6）. 投掷每个筹码使得它落地时等可能地显示任意一面，并与(X_i,Y_i)的秩对应，它们等可能地对应于一个正的D_i（符号秩R_i等于秩）或一个负的D_i（符号秩R_i等于其负秩）. 令T^+是投掷所有n个筹码之后显示正的编号的和，这正好对应于（3）式中T^+的定义. 在筹码游戏中T^+的概率分布与当H_0为真时由（3）式
357 定义T^+的概率分布一样，但是对筹码游戏，它更容易想像.

筹码游戏的样本空间由诸如（1,2,3,-4,-5,6,7,…,n）这样的点组成，像例5.7.1中的数据集一样，仅仅是 R_i 的重排. 由于投掷之间是相互独立的，所以 2^n 个点中每一个出现的概率都是 $(1/2)^n$. 检验统计量 T^+ 为样本点中正数的和. 因此 T^+ 等于任意数 x 的概率就是将正数和为 x 的那些样本点的个数，乘以概率 $(1/2)^n$ 得到.

例如，如果 $n=8$，那么 T^+ 等于0，只有一种方式（即所有的正编号的面都朝下)，所以 $P(T=0)=(1/2)^8$. $T^+=1$ 只有一种方式，$T^+=2$ 只有一种方式，但是 $T^+=3$ 有2种方式，点（-1,-2,3,-4,-5,-6,-7,-8）和（1,2,-3,-4,-5,-6,-7,-8）. $T^+=4$ 也有2种方式. 即

$$P(T^+=0)=(1/2)^8=1/256 \qquad P(T^+\leqslant 0)=0.0039$$
$$P(T^+=1)=1/256 \qquad P(T^+\leqslant 1)=0.0078$$
$$P(T^+=2)=1/256 \qquad P(T^+\leqslant 2)=0.0117$$
$$P(T^+=3)=2/256 \qquad P(T^+\leqslant 3)=0.0195$$
$$P(T^+=4)=2/256 \qquad P(T^+\leqslant 4)=0.0273$$

等　　　　　　　　等

当 $n\leqslant 20$ 时，Owen(1962)把 T^+ 的分布函数列成表格，当 $n\leqslant 50$ 时，T^+ 的分布函数见Harter和Owen(1970)中的表格. 对 $n\leqslant 100$，McCornack(1965)给出了有选择的分位数表，那个表格比我们这里需要的更广泛，所以本书中的表A12给出了更有用的选择分位数表. 用表A12时，一般会导致一个稍微保守的检验，因为小于 p-分位数的概率可能会小于 p. 例如在前一段中，$n=8$，表A12给出了 T^+ 的0.025分位数是4，而对应于 T^+ 小于4的临界域的真正水平是0.0195. Claypool(1970)以及Chow和Hodges(1975)给出了关于 T^+ 精确分布的进一步结果.

对于单边检验，当中位数之差是0时，样本点属于临界域的概率达到最大值，所以这是要考虑的情况. 因此在 H_0 为真的单边检验中，前面 T^+ 的分布是相当有效的.

为了求当数据有结时 T^+ 的条件分布，只是改变上述讨论中最初的一步. 即筹码上的编号必须与赋给所考虑数据集中数对 (X_i,Y_i) 的秩和平均秩一致. 记这些秩与平均秩为 $a_1,a_2,\cdots,a_n$. 在例5.7.1中，我们有，$a_1=1.5,a_2=1.5,a_3=3$，等等. 对这些数，我们可以找到 T^+ 的分布. 因为在例5.7.1中有11个数，在样本空间中有 $2^{11}=2048$ 个点. 这些点中最小的5%，大约是102个点，它们构成临界域. 对于手工制表，这些点的个数太多了，所以用正态逼近. 358

为了用正态逼近，令 S 为所有 R_i 的和. 然后用1.5节中的中心极限定理，我们需要当 H_0 为真时，S 的均值和方差. 注意，在 H_0 下

$$P(R_i=a_i)=1/2 \qquad 和 \qquad P(R_i=-a_i)=1/2$$

所以

$$E(R_i)=a_i(\tfrac{1}{2})+(-a_i)(\tfrac{1}{2})=0 \tag{11}$$

和

$$\mathrm{Var}\,(R_i)=a_i^2\,(\tfrac{1}{2})+(-a_i)^2(\tfrac{1}{2})=a_i^2 \tag{12}$$

因为$\{R_i\}$相互独立（筹码是独立投掷的），我们可以应用定理1.4.1和1.4.3，得到

$$E(S)=\sum_{i=1}^{n}E(R_i)=0 \tag{13}$$

及

$$\mathrm{Var}\,(S)=\sum_{i=1}^{n}\mathrm{Var}\,(R_i)=\sum_{i=1}^{n}a_i^2 \tag{14}$$

但是，由于 a_i^2 总是等于 R_i^2（符号总是 +），我们说

$$\mathrm{Var}\,(S)=\sum_{i=1}^{n}R_i^2 \tag{15}$$

并将中心极限定理应用于

$$T=\frac{\sum_{i=1}^{n}R_i}{\sqrt{\sum_{i=1}^{n}R_i^2}} \tag{16}$$

只要精确表不可用时，就用连续性修正时的正态分布作为近似. Vorlickova(1970)和Conover(1973a)给出了处理结的验证. □

这里提出的处理零差的方法是Wilcoxon(1949)所建议的. Pratt(1959)完整地讨
359 论了另一个处理零差的方法，它包括将零差保留下来，如所描述的对$|D_i|$排序，并将所有的 $D_i=0$ 当作一个结来处理，用通常的方法给它们赋以平均秩. 除了当 $D_i=0$ 时，$R_i=0$ 外，R_i 和通常一样定义，然后由（5）式来计算 T，并与表A1比较. 当检验假设时，不在Pratt的方法中使用表A12，但是，Rahe(1974)给出了一些精确表. Conover(1973b)给出的比较表明，每个处理零点的结的方法都会在某些情况下比其他方法更有效，所以，相比之下没有理由更喜欢某一个. Pratt保留零差的建议将用在下面的来寻找 D_i 的共同的中位数 $d_{0.50}$ 置信区间的方法中，参见Tukey(1949)和Walker和Lev(1953).

▶中位数差的置信区间

数据 数据由 n 个二维随机变量 $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$ 各自的观测值 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 组成. 对每一数对，计算差

$$D_i=Y_i-X_i$$

对每个对，将它们从最小（最负的）到最大的（最正的）排列，记为如下.

$$D^{(1)}\leqslant D^{(2)}\leqslant\cdot\cdot\cdot\leqslant D^{(n-1)}\leqslant D^{(n)}$$

或者在一样本情形中，数据由单样本 $D_1,D_2,\cdots,D_n$ 组成，如所示的次序排列. 我们希望找到 D_i 共同的中位数（或均值）的置信区间.

假定条件

1. 每个 D_i 的分布是对称的.

2. D_i 相互独立.
3. D_i 都有相同的中位数.
4. D_i 的度量尺度至少是区间的.

方法 为得到 $1-\alpha$ 置信区间，从表 A12 中获得 $\alpha/2$ 分位数 $w_{\alpha/2}$.（如果 $w_{\alpha/2}=0$，则对这个 α 值，得不到置信区间.）然后对所有的 i 和 j，考虑 $n(n+1)/2$ 个可能的平均 $(D_i+D_j)/2$，包括 $i=j$，它是 D_i 和它自己的平均，即 D_i. 这些平均中，第 $w_{\alpha/2}$ 大的与第 $w_{\alpha/2}$ 小的构成 $1-\alpha$ 置信区间的上下界. 不必计算 $n(n+1)/2$ 个平均；只须计算最 360
大的和最小的附近的平均来得到置信区间.

计算机辅助 *Minitab* 和 *StatXact* 可给出均值差或中位数差的置信区间，以及著名的 Hodges-Lehmann 位置估计. ◀

例 5.7.1（续）

D_i 的 12 个值按顺序排列为

$$-15, -12, -7, -5, -4, -1, -1, 0, 2, 5, 6, 9$$

要找出中位数差的 95% 置信区间，用 $n=12$，在表 A12 中得到 $w_{0.025}=14$. 这 14 个最小的平均，由 $(-15-15)/2$ 开始，是

$$-15, -13.5, -12, -11, -10, -9.5, -9.5, -8.5, -8, -8, -8, -7.5, -7, -6.5$$

所以，置信区间的下界是 -6.5. 14 个最大的平均数是

$$9, 7.5, 7, 6, 5.5, 5.5, 5, 4.5, 4, 4, 4, 3.5, 3, 2.5$$

所以，置信区间的上界是 2.5. 进取性得分的中位数差（第一个出生的双胞胎进取性得分的中位数减去第二个双胞胎进取性得分的中位数）的 95% 置信区间是

$$P(-6.5 \leqslant d_{0.50} \leqslant 2.5) \geqslant 0.95 \qquad (17)$$

求 14 个最小和最大平均的一个方便的方法是，形成一个平均数的上三角矩阵，其中矩阵的行和列是所有的 D.

	-15	-12	-7	-5	-4	-1	-1	0	2	5	6	9
-15	-15	-13.5	-11	-10	-9.5	-8	-8	-7.5	-6.5	-5	-4.5	-3
-12		-12	-9.5	-8.5	-8	-6.5	-6.5	-6	-5	-3.5	-3	-1.5
-7			-7	-6	-5.5	-4	-4	-3.5	-2.5	-1	-0.5	1
-5				-5	-4.5	-3	-3	-2.5	-1.5	0	0.5	2
-4					-4	-2.5	-2.5	-2	-1	0.5	1	2.5
-1						-1	-1	-0.05	0.5	2	2.5	4
-1							-1	-0.05	0.5	2	2.5	4
0								0	1	2.5	3	4.5
2									2	3.5	4	5.5
5										5	5.5	7
6											6	7.5
9												9

■ 361

□**理论**　为了看差平均 $(D_i+D_j)/2$ 和差的秩之间的关系，考虑如下. 假设没有结，任何 D_i 的秩，如前面的例子中，$D_i=6$ 等于与 $D_i=6$ 相比离0一样近或更近的 D_i 的个数，通过计数包含 D_i 的差平均在0到6之间的数目，我们可得到 D_i 的秩（必须小心在这个计数中，包括 D_i 本身的平均）. 对所有正的 D_i 重复这个步骤，作为总计数，我们得到检验统计量 T^+.

D_i 的中位数 $d_{0.50}$ 的置信区间可以用 Wilcoxon 检验，对各种不同的 m 值检验

$$H_0\colon d_{0.50} = m$$

来求得. 这个方法等价于从每个 D_i 中减去 m 的值，并检验看这个新 D_i 的中位数是否等于零. 但是我们可不用从每个 D_i 中减去 m 的值，然后重新排序并重新计算 T^+，容易看到原来 D_i 的两两平均，数有多少个平均在 m（像我们前面做的，它代替零,）之上，并且其值等于 T^+. 再逆回去，从 T^+ 的临界值开始，找那些最大的平均，停止点是 m 的值，它不能导致接受 H_0. 这样，置信区域的界就找到了. □

Noether(1967b)证明了如果连续性假设不成立，带有端点的置信区间（U 和 L）的置信系数至少是 $1-\alpha$，而没有端点的置信区间（U 和 L）的置信系数至多是 $1-\alpha$. 因此，我们推荐包含端点的置信区间及下列形式的表述：

$$P(L \leqslant d_{0.50} \leqslant U) \geqslant 1-\alpha$$

Moses(1965)给出了这个寻找置信区间方法的一个讨论. 如果抽样是分层的而不是随机的，则参见 McCarthy(1965)的文章. 对于样本中有其他类型的相关性，参见 Høyland(1968). Puri 和 Sen(1968)给出了多元随机变量情形下的置信区域. Geertsema(1970)以及 Srivastava 和 Sen(1973)的文章表明，序贯抽样方法能提供一些优点. Schuster 和 Navarte(1973), Noether(1973), Johns(1974)以及 Maritz, Wu 和 Staudte(1977)讨论了其他估计分布中心的方法. 稳健位置估计量的理论上的讨论参见 Seran(1977).

与其他方法比较

当碰到配对观测，并希望检验均值差是否为零，且度量尺度如本节所述的是区
362 间尺度时，第一个想起的检验通常是“配对 t 检验”，也称作“一样本 t 检验”. 这个检验用检验统计量

$$t = \frac{\overline{D}}{\sqrt{\dfrac{1}{n(n-1)}\sum_{i=1}^{n}(D_i-\overline{D})^2}} \tag{18}$$

其中，$\overline{D}$ 是 D_i 的样本均值，将这个 t 值与表 A12 中第 $k=n-1$ 行中的 t 分布的分位数比较. 为了让表 A12 中的分位数更精确，必须作另外的正态性假设. 即对 Wilcoxon 检验加上假设：D_i 是独立同分布的正态随机变量.

Wilcoxon 检验的假设比正态性的假设更容易验证. 如果数据是离散型的，我们立刻知道分布不是正态的，因为正态分布是连续的. 如果偶尔有很大或很小的观测值，称作“离群值”，则 t 检验的功效大大下降，不应该用它.

如果有 t 检验的计算机程序，则它也可以用于 Wilcoxon 检验. 计算 t 值时，仅用 R_i 代替 D_i，并将结果和表 A21 比较（如刚才描述的）. 这个近似比前面描述的正态逼近稍微精确些，而且对有结的情况也有很好的表现. 细节可参见 Iman (1974a).

Wilcoxon 符号秩检验相对于配对 t 检验的渐近相对效率（A. R. E.）可在下面的约束下进行计算.

1. 二维随机变量 $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ 构成随机样本.

2. 除了均值上的差异外，X_i 与 Y_i 同分布.

在这些条件下，A. R. E. 的范围可能从 $108/125 = 0.864$ 直到无穷，但是，惊奇的是它能保证从不会小于 0.864. 因此 Wilcoxon 检验不会太差，但是在某些条件下，它与通常的参数检验相比可以相当地好.

更进一步地说，如果差 D_i 有正态分布，则 A. R. E. 是 $3/\pi = 0.955$. 如果我们假设差 D_i 的分布换为均匀分布，则 A. R. E. 是 1.0. 对双指数分布来说，则 A. R. E. 是 1.5.

在前面的约束下，符号检验（3.4 节）可以用于检验和 Wilcoxon 检验相同的假设. 那么这时，符号检验关于 Wilcoxon 检验有如下的 A. R. E.： 363

假设的分布	A.R.E.
正态	$\frac{2}{3}$
均匀	$\frac{1}{3}$
双指数	$\frac{4}{3}$

令人惊讶的是，在有些情况下，符号检验比 Wilcoxon 检验更有效. 与配对 t 检验相比，Wilcoxon 的 A. R. E. 对于双指数分布是 3/2，因此，两个 A. R. E. 相乘就给出符号检验关于配对 t 检验的 A. R. E.，它为

$$\left(\tfrac{4}{3}\right)\left(\tfrac{3}{2}\right) = 2$$

对双指数分布，符号检验的渐近效率是配对 t 检验的 2 倍. 然而在这个情形下，A. R. E. 不是小样本效率好的指标，见 Conover, et al. (1978) 对两样本情形的讨论. 对于来自双指数分布的小样本，Wilcoxon 符号秩检验的效率比符号检验和 t 检验都好. 对于其他的功效和效率的研究，见 Klotz (1963, 1965), Arnold (1965), Noether (1967a) 和 Kraft 和 van Eeden (1972).

Wilcoxon 符号秩检验有时也叫做对称性检验. Schuster (1975) 以及 Rao, Schuster 和 Littell (1975) 讨论了对称分布的估计，同时 Rothman 和 Woodroofe (1972) 提出了另外的对称性检验. Bell 和 Haller (1969), Hollander (1971) 以及 Bhattacharyya, Johnson 和 Neave (1971) 讨论了二维随机变量的对称性检验. Bennett (1965) 以及 Sen 和 Puri (1967) 研究了推广到多元随机变量的情形. Hollander (1970) 将 Wilcoxon 检验用于检验两条回归线的平行. Miller (1970), Weed, Bradley 和 Govindarajulu (1974), Sen 和 Ghosh (1974), Reynolds (1975) 以及 Spurrier 和 Hewett (1976) 提出了适应性序贯抽样方法. 其他和本节相关的文章见 Groeneveld (1972) 以及 Bickel 和 Lehmann (1975).

Walsh(1949)提出的检验和 $n<7$ 的 Wilcoxon 检验相同，但不同于 n 大于等于 7 的情形. Schach(1969a)给出了由 Batschelet(1965)提出的 Wilcoxon 检验包括圆分布在内的应用.

习题

1. 选择由 20 个开机动车的司机组成一个随机样本，看酒精是否影响反应时间. 在实验室中测量了每个司机在喝一定量含有酒精的饮料前后的反应时间. 这反应时间如下
364 (以秒计).

对象	前	后	对象	前	后
1	.68	.73	11	.65	.72
2	.64	.62	12	.59	.60
3	.68	.66	13	.78	.78
4	.82	.92	14	.67	.66
5	.58	.68	15	.65	.68
6	.80	.87	16	.76	.77
7	.72	.77	17	.61	.72
8	.65	.70	18	.86	.86
9	.84	.88	19	.74	.72
10	.73	.79	20	.88	.97

酒精影响反应时间吗?

2. 一名食品店主希望看到，是否可以认为每次出售时顾客所买物品数量的中位数是 10，所以他在收款台前观察了 12 名顾客.

顾客	物品数量	顾客	物品数量
1	22	7	15
2	9	8	26
3	4	9	47
4	5	10	8
5	1	11	31
6	16	12	7

可以用 Wilcoxon 检验吗? 在这个问题中，违反了模型的哪个假设?

3. 检验例 3.5.3 的数据，看第二年的观测值是否倾向于比第一年的观测值小.

4. 给女子篮球队的每名成员一个简短的热身，然后让她们每人罚 25 个球，记录其命中的个数 X. 然后给这个队以大运动量训练，在简短的休息后，让她们每人再罚 25 个球，再记录其命中的个数 Y. 这些数据表明当运动员累了时罚球命中率会下降吗?

	运动员											
	1	2	3	4	5	6	7	8	9	10	11	12
X_i(之前)	18	12	7	21	19	14	8	11	19	16	8	11
Y_i(之后)	16	10	8	23	13	10	8	13	9	8	8	5

5. 参与竞选的候选人意识到，如果她挑选她所在部门的中间职位，则她的得票率最大. 因此，她设计了一个问卷，将其发给 15 个投票者（类似于随机样本）. 问卷结果得分从一
365 个极端值（0）打到另一个极端值（10）.

投票者	分数	投票者	分数	投票者	分数
1	6.7	6	9.3	11	8.8
2	4.2	7	8.9	12	5.4
3	4.1	8	7.4	13	6.1
4	2.3	9	7.4	14	6.0
5	6.1	10	9.3	15	4.9

求这个得分中位数的90%置信区间. 基于本节的方法，你的得分中位数的点估计是什么?

6. 急救班负责一个又长又窄的湖上的安全，他们希望在现场建一个永久站点，可以最小化将来他们到达事故现场的总距离. 那个地方应该在事故发生地点的中间位置. 假设已发生的事故类似于所有今后可能发生的事故的随机样本，测量（到水坝）的距离如下.

事故	距离（英里）	事故	距离（英里）
1	7.1	8	6.1
2	4.4	9	2.2
3	3.9	10	6.7
4	2.2	11	4.9
5	4.2	12	7.3
6	3.4	13	0.3
7	1.1	14	7.6

水坝到站点的最佳距离的95%置信区间是什么?

7. 随机选到7对已婚夫妇，问每位丈夫及妻子今年会花多少钱来给配偶买圣诞礼物. 回答如下.

	夫妇						
	1	2	3	4	5	6	7
丈夫	25	21	38	64	52	16	26
妻子	16	42	56	41	19	26	24

(a) 求出丈夫所花钱数超过妻子所花钱数的中位数的95%置信区间.

(b) 你求得的区间的精确置信水平是多少?

8. 4名报考研究生的学生参加了两次GMAT，成绩如下.

学生	第一次考试	第二次考试
1	470	510
2	530	550
3	610	600
4	440	490

366

(a) 求出Wilcoxon符号秩检验统计量的精确分布，即正符号秩的和，并且画出它的分布函数图.

(b) 在(a)部分所作的图中，标出检验统计量值的位置，无论哪个更小，求出右边检验的精确 p-值，或者是左边检验的 p-值.

(c) 对于平均得分增量，求出它非参数的80%（近似地）置信区间.

(d) 所求得的区间的精确置信水平是多少?

思考题

1. 当 $n=5$ 时，求本节中检验统计量 T^+ 的概率分布（假设双边检验中 H_0 为真）.

2. 在 Wilcoxon 符号秩检验中，不考虑零差以便使用精确表．为什么当我们寻找中位数差的置信区间时最好考虑零差呢？
3. 如果我们用符号秩 R_i 来计算 t 统计量（（18）式）而不是使用差 D_i．证明这个统计量就是如下 T 的函数．

$$t_R = \frac{T}{\left(\frac{n}{n-1} - \frac{1}{n-1}T^2\right)^{\frac{1}{2}}}$$

正如（5.12.3）式所示．并且证明：当 T 上升时，t_R 也上升．因此当 T 足够大时拒绝 H_0，等价于 t_R 足够大时拒绝 H_0．
4. 用习题 4 中的数据来计算配对 t 检验统计量，并与 Wilcoxon 检验相比较．（用表 A21，其中行 $k=11$；在配对 t 检验中要考虑零．）

5.8 多个相关的样本

在 5.2 节中，我们介绍了关于多个独立样本的 Kruskal-Wallis 秩检验，这也是对于 5.1 节中所介绍的两个独立样本 Mann-Whitney 检验的拓广．在这一节，我们将考虑分析多个相关样本的问题，这也是我们在前一节中所考虑的配对或两个相关样本问题的拓广．首先，我们将提出 Friedman 检验，它是 3.4 和 3.5 节中所讲述符号检验的拓广，然后我们将介绍 Quade 检验，即前一节中介绍的 Wilcoxon 符号秩检验的拓广．在这两个检验中，Friedman 检验更有名，且需要的假定比较少，但是当只有 3
367 个处理时，它缺乏功效，就像只有 2 个处理时，符号检验的功效不如 Wilcoxon 符号秩检验一样．当有 4 个或 5 个处理时，Friedman 检验的功效和 Quade 检验的几乎一样．但是当处理的个数是 6 个或 6 个以上时，Friedman 检验会有更高的功效．功效和 A. R. E. 的比较可参见 Iman et al.(1984)以及 Hora 和 Iman(1988)．

设计一个用来检测 $k(k\geq 2)$ 个可能不同处理中的差异的试验，可以引出多个相关样本的问题．在区组内排列观测，这些区组是 k 个在某些重要方面相似的试验单位的组，如 k 个同窝出生的仔畜，那么这些仔畜对某种特殊刺激的反应较之在任意的窝里随机选取仔畜的反应更加相似．对这一个组内的 k 个试验单元随机配以 k 种仔细的处理，所以每种处理在每个组内只执行一次．以这种方式，对处理之间进行相互比较，且没有过多的混淆试验结果的影响．记 b 为所用区组的总数，$b>1$．

这里描述的试验排列通常叫做随机化的完全区组设计（randomized complete block design）．这个设计可以与下一节描述的不完全（incomplete）区组设计进行比较，而不完全区组设计不包含足够的试验单元，使所有的处理应用到所有区组，所以，每个处理出现在一些区组中，但不会出现在其他区组中．随机化的完全区组设计的例子如下．

1. 心理学（Psychology）．共 5 窝老鼠，每窝中有 4 只老鼠，用于考查环境与侵犯性之间的关系．我们认为每一窝是一个区组．设计了 4 组不同的环境．把每窝中的一只老鼠放在一个环境中，使得从每窝中取出的 4 只老鼠在 4 个不同的环境中．在

一定长的时间后，对老鼠和它们的同窝出生仔畜重新分组，且根据侵犯度进行排序.

2. 家庭经济（Home economics）. 比较6种不同类型的生面包团，每一类型的生面包团形成3块，看哪个烤得最快. 用3种不同的烤箱，且每个烤箱同时烤6种不同类型的面包. 烤箱是区组，生面包团是处理

3. 环境工程（Environmental engineering）. 如果不同的处理可以应用到同一个单元而没有留下残余影响，则一个试验单元可以形成一个区组. 7个不同的男士参与配色方案对工作效率影响的研究. 认为每个人是一区组，且在3个房间的每一间都待一段时间，每一间都有自身类型的色彩设计. 在房间里时，每个人只执行一个任务，并度量其工作效率. 3个房间则是处理. 368

到目前为止，读者应该对随机化的完全区组设计的种类有了一些概念. 检验零假设为没有处理区别的通常参数方法叫做双因素方差分析. 下面的非参数方法只依赖于每个区组内观测的秩. 因此，可以考虑用秩的双因素方差分析. 这个检验将按它的发明者，一个著名的经济学家 Milton Friedman 来命名.

▶Friedman 检验

数据 数据是由 b 个相互独立的 k 维随机变量组成，记为 $(X_{i1}, X_{i2}, \cdots, X_{ik})$，其中 $i = 1, 2, \cdots, b$，称为 b 个区组. 随机变量 X_{ij} 表示在第 i 区组中用处理 j 的样本. b 个区组排列如下：

	处理			
区组	**1**	**2**	···	k
1	X_{11}	X_{12}	···	X_{1k}
2	X_{21}	X_{22}	···	X_{2k}
3	X_{31}	X_{32}	···	X_{3k}
···	···	···	···	···
b	X_{b1}	X_{b2}	···	X_{bk}

记 $R(X_{ij})$ 为秩，它从1到 k，赋给区组（行）i 中的 X_{ij}，即对于区组 i，相互比较随机变量 $X_{i1}, X_{i2}, \cdots, X_{ik}$，最小观测值赋以秩1，第二小的观测值赋以秩2，依此类推，区组 i 中的最大观测值赋以秩 k. 给所有 b 个区组进行赋秩. 如果有结，则用平均秩.

我们对每一处理的秩求和得到 R_j，其中，对 $j = 1, 2, \cdots, k$

$$R_j = \sum_{i=1}^{b} R(X_{ij}) \tag{1}$$

假定条件

1. b 个 k 维随机变量是相互独立的（一个区组的结果不会影响其他区组的结果）.

2. 每个区组内的观测可以根据某些感兴趣的准则进行排序. 369

检验统计量 Friedman 建议用如下统计量：

$$T_1 = \frac{12}{bk(k+1)} \sum_{j=1}^{k} \left(R_j - \frac{b(k+1)}{2} \right)^2 \tag{2}$$

如果有结现象存在，那么我们需要进行相应的调整．记 A_1 为秩或平均秩的平方和，则有

$$A_1 = \sum_{i=1}^{b}\sum_{j=1}^{k}[R(X_{ij})]^2 \tag{3}$$

"校正因子" C_1 由下式计算

$$C_1 = bk(k+1)^2/4 \tag{4}$$

因为有结出现，经校正后的统计量 T_1 变为

$$T_1 = \frac{(k-1)\left[\sum_{j=1}^{k} R_j^2 - bC_1\right]}{A_1 - C_1} = \frac{(k-1)\sum_{j=1}^{k}\left(R_j - \frac{b(k+1)}{2}\right)^2}{A_1 - C_1} \tag{5}$$

最近的研究表明，在秩 $R(X_{ij})$ 上计算出的双因素方差分析的统计量 T_2，由于有更精确的逼近分布，因此受到人们的喜欢．T_2 可简化成上面所给出的 T_1 的函数．

$$T_2 = \frac{(b-1)T_1}{b(k-1) - T_1} \tag{6}$$

这些逼近表达的细节参见 Iman 和 Davenport(1980)．

零分布　我们很难找到统计量 T_1（或 T_2）的精确分布，因此我们往往使用它们的逼近分布．T_1 的逼近分布是自由度为 $k-1$ 的 χ^2 分布．但是，有时候这个逼近分布的近似程度并不好，因此我们推荐用 T_2 而不用 T_1，当零假设成立时，它的近似分位数由自由度为 $k_1 = k-1, k_2 = (b-1)(k-1)$ 的 F 分布给出（见表 A22）．

假设

H_0:同一个区组中，对随机变量的每个赋秩是等可能的(即处理效果相同)．

370 H_1:至少有一个处理倾向于比其他处理中的至少一个处理产生较大的观测值．

如果统计量 T_2 大于 F 分布的 $1-\alpha$ 分位数，则我们以 α 近似水平拒绝 H_0，其中，F 分布的自由度为 $k_1 = k-1, k_2 = (b-1)(k-1)$．这个近似结果相当好，并且随着 b 的增大近似效果越好．我们可以根据表 A22 估计相应的近似 p-值．

多重比较　只有当 Friedman 检验导致拒绝零假设时，我们可用下述方法来比较各个处理．如果下列不等式满足，则认为处理 i 和 j 是不相同的．

$$|R_j - R_i| > t_{1-\alpha/2}\left[\frac{2(bA_1 - \sum R_j^2)}{(b-1)(k-1)}\right]^{\frac{1}{2}} \tag{7}$$

其中，R_i, R_j 和 A_1 如上所述，$t_{1-\alpha/2}$表示自由度为（$(b-1)(k-1)$）的 t 分布的上侧 $1-\alpha/2$分位数，α 值与上面 Friedman 检验所用的值相同．

另外，(7) 式也可以表示为关于统计量 T_1 的函数，

$$|R_j - R_i| > t_{1-\alpha/2}\left[\frac{(A_1 - C_1)2b}{(b-1)(k-1)}\left(1 - \frac{T_1}{b(k-1)}\right)\right]^{\frac{1}{2}} \tag{8}$$

如果没有结存在，则 (7) 式中的 A_1 可以化简为：

$$A_1 = bk(k+1)(2k+1)/6$$

同时，(8) 式中的 $(A_1 - C_1)$ 可以化简为：

$$A_1 - C_1 = bk(k+1)(k-1)/12$$

计算机辅助 在 *Minitab*, *Splus*, *SAS* 和 *StatXact* 中都有 Friedman 检验的计算机程序. ◀

例 5.8.1

随机选出 12 名私房业主参与一个种植苗圃的试验. 要求每一个业主从他的院子里选出合理而等面积的 4 块地种植 4 种不同的草，每一块地种一种. 在规定的时间段结束后，要求房主根据加权一些重要的指标如支出、养护要求、漂亮程度、耐寒性、妻子的喜好等等，将 4 种草排序，其中秩 1 赋给最不受喜爱的一种，秩 4 赋给最受喜爱的一种. 零假设为：4 种草的受喜爱程度没有区别，备择假设为：有某种草比另一些草更受偏爱. 把 12 个区组的每一组都平均分成了等面积的 4 块，受到了基本相同的照料，因为假设了这 4 块都由同一个私房业主照料. 试验的结果如下： 371

	草			
房主	1	2	3	4
1	4	3	2	1
2	4	2	3	1
3	3	1.5	1.5	4
4	3	1	2	4
5	4	2	1	3
6	2	2	2	4
7	1	3	2	4
8	2	4	1	3
9	3.5	1	2	3.5
10	4	1	3	2
11	4	2	3	1
12	3.5	1	2	3.5
R_j（总和）	38	23.5	24.5	34

首先，$A_1 = 356.5$，它是所有 $R(X_{ij})$ 的平方和，即总平方和. (4) 式给出了

$$C_1 = \frac{12(4)(25)}{4} = 300$$

(5) 式给出了

$$T_1 = \frac{3[(38)^2 + (23.5)^2 + (24.5)^2 + (34)^2 - 12(300)]}{356.5 - 300}$$
$$= 8.097$$

把 T_1 带入 (6) 式，得到

$$T_2 = \frac{11(8.097)}{12(3) - 8.097} = 3.19$$

近似水平为 $\alpha = 0.05$ 的临界域对应着所有大于 2.90 的 T_2 值. 根据表 A22，我们得到 $k_1 = 3, k_2 = 33$ 的 F 分布的 0.95 分位数是 2.9，即如果 $T_2 > 2.9$，则拒绝零假设. 因此我们拒绝零假设，并可以得出结论，某种草比另一些草更受偏爱. p-值约为

0.04，它由对表A22中的值做内插得到. 这意味着小到显著水平 $\alpha=0.04$，我们也
372 能拒绝零假设. 对于多重比较，根据表A21得到自由度为（11）（3）=33的 t 分布的分位数 $t_{0.975}$ 为2.036. 由（7）式，我们得到

$$t_{0.975}\left[\frac{2(bA_1-\sum R_j^2)}{(b-1)(k-1)}\right]^{\frac{1}{2}}=11.49$$

对任意的2种草，如果它们的秩和大于11.49个单位，那么认为它们是不平等的. 因此认为第1种草要优于第2种和第3种草，其他的区别则不显著. ■

下面的检验方法同样也是检验在随机化的完全区组设计中，两种处理均值相等的零假设，但是它利用了有关的区组极差的信息，对于极差较大的区组赋予更大的权重. 所应用区组内赋秩的方法与Friedman检验方法相同. 我们将根据它的发明者，Dana Quade(1972,1979)来命名这个检验.

▶Quade检验

数据 先按照前一个检验所描述的赋秩方法给区组内的 $R(X_{ij})$ 赋秩，下一步又一次用到了原始的观测 X_{ij}，然后根据每一区组极差的大小，给每一个区组自身赋秩，其中区组的极差定义为最大和最小观测值的差值.

$$\text{区组 } i \text{ 的极差}=\underset{j}{\text{maximum}}\{X_{ij}\}-\underset{j}{\text{minimum}}\{X_{ij}\} \tag{9}$$

每一个区组计算得到一个样本极差，共有 b 个样本极差，其中极差最小的区组赋秩1，第二小的区组赋秩2，依此类推，极差最大的区组赋秩为 b. 如果有结存在，则使用平均秩. 记 $1,2,\cdots,b$ 区组各组的秩分别为 $Q_1,Q_2,\cdots,Q_b$.

最后，用区组的秩 Q_i 乘以区组中的秩 $R(X_{ij})$ 与区组内的平均秩 $(k+1)/2$ 的差值，得到乘积 S_{ij}，其中

$$S_{ij}=Q_i\left[R(X_{ij})-\frac{k+1}{2}\right] \tag{10}$$

373 是代表区组内每一个观测的相对大小的统计量，调整使之反映区组出现的相对显著性.

记 S_j 为第 j 个处理 S_{ij} 的和：

$$S_j=\sum_{i=1}^{b}S_{ij} \tag{11}$$

其中，$j=1,2,\cdots,k$.

假定条件 前两个假设与前面检验中提出的假设相同. 由于我们在检验中需要对区组进行比较，所以需要如下的第3个假设.

3. 每一个区组可以确定样本极差，因此，可以给区组赋秩.

检验统计量 为了简便，首先计算项

$$A_2=\sum_{i=1}^{b}\sum_{j=1}^{k}S_{ij}^{2} \tag{12}$$

其中，S_{ij}由（10）式给出，A_2 称为“总平方和”. 如果没有结存在，则 A_2 可以化简为

$$A_2 = b(b+1)(2b+1)k(k+1)(k-1)/72 \tag{13}$$

然后计算项

$$B = \frac{1}{b}\sum_{j=1}^{k} S_j^2 \tag{14}$$

其中，S_j 由（11）式给出，B 称为“处理平方和”. 检验统计量是

$$T_3 = \frac{(b-1)B}{A_2 - B} \tag{15}$$

如果 $A_2 = B$，则认为点在临界域内，并可以求得 p-值为 $(1/k!)^{b-1}$.

注意，T_3 是一个由（10）式所给出的得分 S_{ij}计算出的双因素方差分析检验统计量.

零分布 T_3 的精确分布很难得到，因此，如同前面 Friedman 检验中的那样，可用自由度分别为 $k_1 = k-1, k_2 = (b-1)(k-1)$ 的 F 分布来逼近，表 A22 给出了 F 分布的分位数. 374

假设 假设与 Friedman 检验的假设相同.

当 T_3 超过自由度为 $k_1 = k-1, k_2 = (b-1)(k-1)$ 的 F 分布的 $1-\alpha$ 分位数（由表 A22 获得），则我们在 α 的水平上拒绝零假设. 事实上，F 分布只是 T_3 精确分布的一个近似分布，而精确分布表现在我们还无法求得. 随着 b 值的增大，F 近似分布将接近于更精确.

多重比较 只有当前面的检验结果为拒绝零假设时，我们才可做多重比较. 如果以下不等式成立，

$$|S_i - S_j| > t_{1-\alpha/2}\left[\frac{2b(A_2 - B)}{(b-1)(k-1)}\right]^{1/2} \tag{16}$$

我们则认为处理 i 和 j 不相同，其中，S_i, S_j, A_2, B 如前面说述，$t_{1-\alpha/2}$是自由度为 $(b-1)(k-1)$ 的 t 分布的 $1-\alpha/2$ 分位数，它由表 A21 获得. 对所有的处理对进行这样的比较，α 与 Quade 检验中所用的相同.

计算机辅助 *StatXact* 包含有 Quade 检验的计算机程序.

例 5.8.2

抽取 7 个商店进行市场调查，在每一个商店中，5 种不同品牌的新型洗手液依次排开. 在一周结束的时候，计算每种品牌的销售瓶数，结果列表如下：

顾客人数（商店内部的秩）

商店	品牌 A	B	C	D	E
1	5 (2)	4 (1)	7 (3)	10 (4)	12 (5)
2	1 (2.5)	3 (5)	1 (2.5)	0 (1)	2 (4)
3	16 (2)	12 (1)	22 (3.5)	22 (3.5)	35 (5)
4	5 (4.5)	4 (2.5)	3 (1)	5 (4.5)	4 (2.5)
5	10 (3.5)	9 (2)	7 (1)	13 (5)	10 (3.5)
6	19 (2)	18 (1)	28 (3)	37 (4)	58 (5)
7	10 (5)	7 (2.5)	6 (1)	8 (4)	7 (2.5)

每一个商店内把每一种品牌赋秩于 1 到 5，如果存在结，则用平均秩，括号中的值为秩 $R(X_{ij})$.

接着，计算每一个商店内的样本极差，即最大观测值减去最小观测值. 在商店 1
375 中，样本的极差是 $12-4=8$. 这些样本极差，样本极差的秩 Q_i 以及乘积如下表所列，

$$S_{ij}=Q_i[R(X_{ij})-(k+1)/2]$$
$$S_{ij}=Q_i[R(X_{ij})-3]$$

商店数	样本极差	秩 Q_i	品牌 A	B	C	D	E
1	8	5	−5	−10	0	+5	+10
2	3	2	−1	+4	−1	−4	+2
3	23	6	−6	−12	+3	+3	+12
4	2	1	+1.5	−0.5	−2	+1.5	−0.5
5	6	4	+2	−4	−8	+8	+2
6	40	7	−7	−14	0	+7	+14
7	4	3	+6	−1.5	−6	+3	−1.5
		$S_j=$	−9.5	−38	−14	+23.5	+38

由（12）式得，

$$A_2=\sum_{i=1}^{7}\sum_{j=1}^{5}S_{ij}^2=(-5)^2+(-10)^2+\cdots=1366.5$$

略小于由（13）式对没有结情况下得到的 1400. 由（14）式，得到

$$B=\frac{1}{7}\sum_{j=1}^{5}S_j^2=\frac{1}{7}[(-9.5)^2+(-38)^2+\cdots]=532.4$$

再带入（15）式，就得到检验统计量

$$T_3=\frac{6(532.4)}{1366.5-532.4}=3.83$$

T_3 值大于 2.78，它是由表 A22 得出的自由度 $k_1=4, k_2=24$ 的 F 分布的 0.95 分位数，因此当 $\alpha=0.05$ 时，拒绝零假设. 事实上，仔细查看表 A22，可发现 p-值略小于 0.025. 因此我们可以得出结论，某些品牌与其他品牌相比似乎更受顾客欢迎.

因为拒绝了零假设，因此可以运用多重比较. 由（16）式，可以认为两个处理不同，如果它们的 $|S_i-S_j|$ 大于

376
$$t_{1-\alpha/2}\left[\frac{2b(A_2-B)}{(b-1)(k-1)}\right]^{1/2}=2.064\left[\frac{14(834.1)}{24}\right]^{1/2}=45.53$$

其中，$t_{1-\alpha/2}=t_{0.975}$，它是由表 A21 得到的自由度为 $(b-1)(k-1)=24$ 的 t 分布的 0.975 分位数. 因此可以认为品牌 A 和 E，B 和 D，B 和 E 以及 C 和 E 彼此不同.

注意，下图显示出多重比较的结果，其中，字母按照平均得分的升序排列，下划线表示处理没有明显的区别.

B C A D E ■

□**理论** 如果零假设成立，即在一个区组中，每一种赋秩都是等可能的，则我们可以得到 T_1, T_2 和 T_3 的精确分布. 那么在同一个区组中，秩 $R(X_{ij})$ 可能的排列方式有 $k!$ 种，因此在全部 b 个区组排列的数组中有 $(k!)^b$ 种秩的排列方式. 前面提出的假设，表示在零假设成立的前提下，这 $(k!)^b$ 种排列方式出现的可能性是相同的. 因此，给定样本数 k 和区组 b，我们可以通过列出秩的所有可能排列方式，并对每一种排列计算 T_1, T_2, T_3，从而得到它们的概率分布函数.

例如，如果 $k=2, b=3$，则有 $(2!)^3=8$ 种等可能秩的排列方式，下表列出了每一种排列方式相应的 T_1, T_2 值，我们将在后面的内容中考虑 T_3.

	排列							
区组	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1, 2	1, 2	1, 2	2, 1	2, 1	2, 1	1, 2	2, 1
2	1, 2	1, 2	2, 1	1, 2	2, 1	1, 2	2, 1	2, 1
3	1, 2	2, 1	1, 2	1, 2	1, 2	2, 1	2, 1	2, 1
概率	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
T_2 的值	∞	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	∞
T_1 的值	3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	3

因此，当 H_0 成立时，T_1 的概率分布为 $P(T_1=\frac{1}{3})=3/4, P(T_1=3)=1/4$，$T_2$ 的概率分布为 $P(T_2=\frac{1}{4})=3/4, P(T_2=\infty)=1/4$.

为了检查 T_3 在零假设下的行为，如前所述，我们再看一下秩 $R(X_{ij})$ 的 8 种可能的排列情况. 用每一个秩减去平均秩 1.5，此时我们考虑区组的秩分别为 $Q_1=1, Q_2=2, Q_3=3$ 的情况. S_{ij} 的计算结果如下. 377

	排列			
区组	**1**	**2**	**3**	**4**
1	−0.5, +0.5	−0.5, +0.5	−0.5, +0.5	+0.5, −0.5
2	−1, +1	−1, +1	+1, −1	−1, +1
3	−1.5, +1.5	+1.5, −1.5	−1.5, +1.5	−1.5, +1.5
条件概率	1/8	1/8	1/8	1/8
T_3 的值	12	0	$\frac{4}{19}$	$1\frac{3}{13}$

	排列			
区组	**5**	**6**	**7**	**8**
1	+0.5, −0.5	+0.5, −0.5	−0.5, +0.5	+0.5, −0.5
2	+1, −1	−1, +1	+1, −1	+1, −1
3	−1.5, +1.5	+1.5, −1.5	+1.5, −1.5	+1.5, −1.5
条件概率	1/8	1/8	1/8	1/8
T_3 的值	0	$\frac{4}{19}$	$1\frac{3}{13}$	12

T_3 取各个值的概率是：

$$\tfrac{1}{8}\cdot P(Q_1=1,\ Q_2=2,\ Q_3=3)$$

由于1/8表示给定Q_1,Q_2,Q_3的秩时的T_3值的条件概率．假设考虑给Q_1,Q_2,Q_3一个不同的赋秩，如$Q_1=2,Q_2=1,Q_3=3$，那么，如同我们刚才所做的，通过列出S_{ij}值的8个排列，读者很容易再一次验证．我们又看到T_3的8个同样的值，且这8个值每个都有概率

$$\tfrac{1}{8}\cdot P(Q_1=2;\ Q_2=1,\ Q_3=3)$$

计算Q_1,Q_2,Q_3中所有6种（3！）置换排序方式，我们可以得到T_3所取各个值的总概率：1/8．因此，为了计算T_3的零分布，我们这里仅需要考虑一种情形，$Q_i=i,i=1,2,3$．因此通过计数T_3取各个等值的数目，而得到T_3的概率分布

$$P(T_3=0)=1/4,\qquad P(T_3=\tfrac{4}{13})=1/4\qquad P(T_3=1\tfrac{3}{13})=1/4,\qquad P(T_3=12)=1/4$$

根据中心极限定理，我们可以用F分布或χ^2分布作为T_1,T_2,T_3的近似分布．由于有些知识细节超出了本书的范围，因此我们省略掉近似分布的整个推导过程．读者可以参考著作Quade(1972,1979)或Lawler(1978)了解T_3,Iman和Davenport(1979)
378 了解T_2，以及Friedman(1937)了解T_1．□

前面所得到的T_1,T_2,T_3的假设都是基于H_0为真的情形．如果H_0是假的，那么处理和S_j及R_j可能分别与它们的平均值0和$b(k+1)/2$有较大区别，使得3个统计量的值趋于增加．因此我们的判决法则就是：如果T_1,T_2,T_3的值很大，则拒绝零假设H_0．

只有当数据服从方差相同的正态分布时，我们才能用参数方法分析来自随机化的完全区组设计的数据．零假设为：在同一个区组中的随机变量具有相同的均值．检验统计量为

$$F=\frac{(b-1)SSB}{SST-SSB-SSR}\tag{17}$$

其中

$$SSB=\frac{1}{b}\sum_{j=1}^{k}T_j^2-\frac{T^2}{bk}\tag{18}$$

$$SSR=\frac{1}{k}\sum_{i=1}^{b}\left(\sum_{j=1}^{k}X_{ij}\right)^2-\frac{T^2}{bk}\tag{19}$$

$$SST=\sum_{i=1}^{b}\sum_{j=1}^{k}X_{ij}{}^2-\frac{T^2}{bk}\tag{20}$$

$$T_j=\sum_{i=1}^{b}X_{ij}\tag{21}$$

和

$$T=\sum_{i=1}^{b}\sum_{j=1}^{k}X_{ij}\tag{22}$$

并将（17）式中的统计量和表A22中得出的$k_1=k-1,k_2=(b-1)(k-1)$的F分布

的分位数进行比较.

如果我们使用秩 $R(X_{ij})$ 替代数据计算得到这个 F 统计量，则得到的统计量和 T_2 一样；如果使用加权的秩 S_{ij} 替代数据计算得到 F 统计量，则得到统计量 T_3. 对有结的调整也就自动体现在 F 统计量 T_3 中. 我们这里描述的多重比较方法只是一个参数方法，称为 Fisher 最小显著差异方法（LSD），但是，它是通过用 S_{ij} 或者 $R(X_{ij})$ 来计算的，而不是用数据.

对于两个样本（$k=2$），相对于通常的参数 t 检验，Friedman 检验的 A. R. E. 与符号检验相同，即 $2/\pi=0.637$，这里 t 检验是最有功效的检验. 对于 k 个样本的情况，Friedman 检验相对于 F 检验的 A. R. E. 依赖于样本个数 k，如果总体服从正态分布，则 A. R. E. 等于 $0.955k/(k+1)$；如果总体服从均匀分布，则 A. R. E. 等于为 $k/(k+1)$；如果总体服从双指数分布，则 A. R. E. 等于 $3k/2(k+1)$. 在纯漂移型的备 379
择假设下，Friedman 检验相对于常用的 F 检验总不会小于 $0.864k/(k+1)$. Noether (1967a) 完整地讨论了 Friedman 检验的 A. R. E. 问题.

对于 $k=2$，当分布为正态分布时，Quade 检验相对于通常的参数 t 检验的 A. R. E. 与 Wilcoxon 符号秩检验相同，即 $3/\pi=0.955$. 对于 Wilcoxon 检验，它相对于 t 检验的 A. R. E. 从不小于 0.864，但是可能达到无穷. 对于两个样本以上的情形，Quade 检验的 A. R. E. 还未解决，但是可以使用模拟的方法（Iman et al,1984），结果表明，当处理数等于或大于 5 时，Quade 检验不如 Friedman 检验有效.

▶有序备择假设的 Page 检验

在 5.4 节中，对于指定了处理效应顺序的备择假设，我们给出了 k 个独立样本情况下的 Jonckheere-Terpstra 检验，它等价于计算观测值和备择假设中指定处理次序间的 Kendall τ. 顺便提一下，我们也能较好地利用 Spearman ρ.

在随机化的完全区组设计中，Spearman ρ 用于检验 k 个相关样本，其中，备择假设为处理效应服从指定的顺序. Page(1963) 介绍了 Friedman 区组内排序与备择假设 H_1 中指定处理排序间的相关性在检验中的应用.

由于数据中往往有许多结，Page 使用了一个比较简单的统计量，如果在没有结的情况下，它是一个关于 Spearman ρ 的单调函数，即

$$T_4=\sum_{j=1}^{k} jR_j = R_1+2R_2+\cdots+kR_k$$

其中，R_j 表示 Friedman 检验中的处理的秩和，备择假设 H_1 指定处理效应按照升序排列.

虽然 Page(1963) 给出了精确表，但是我们在此仅考虑其大样本逼近，对水平为 α 的右边检验，当

$$T_5=\frac{T_4-bk(k+1)^2/4}{[b(k^3-k)^2/144(k-1)]^{1/2}}$$

超过标准正态分布的 $1-\alpha$ 分位数时（由表 A1），则拒绝 H_0. *StatXact* 可以计算出
380 Page 检验中精确的 p-值.

例 5.8.3

健康研究者推测有规律的运动可以降低人休息时的心率. 为了检验这一理论，8 名平时不经常运动的健康志愿者参加了一个试验，在监督下进行有规律的运动. 在项目开始时测量他们休息时的心跳，此后每一个月测量一次，持续 4 个月.

零假设是：没有差异

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

有序备择假设为：

$$H_1: \mu_1 \leqslant \mu_2 \leqslant \mu_3 \leqslant \mu_4 \leqslant \mu_5$$

其中，μ_1 是 4 个月结束后的均值，μ_5 是初始均值，在 H_1 中，至少有一个严格不等式成立.

观测到的心率以及它们的 Friedman 区组内的秩如下所示

人	初始值	第 1 个月	第 2 个月	第 3 个月	第 4 个月
1	82(4)	84(5)	77(2)	76(1)	79(3)
2	80(4.5)	80(4.5)	76(1.5)	76(1.5)	78(3)
3	75(3)	78(5)	77(4)	74(2)	72(1)
4	65(1.5)	72(5)	68(4)	65(1.5)	66(3)
5	77(5)	74(2)	72(1)	75(3.5)	75(3.5)
6	68(4)	69(5)	65(2)	66(3)	64(1)
7	70(3.5)	74(5)	68(1.5)	70(3.5)	68(1.5)
8	77(4)	76(3)	78(5)	72(2)	70(1)
	$R_5 = 29.5$	$R_4 = 34.5$	$R_3 = 21$	$R_2 = 18$	$R_1 = 17$

注意，R_1 是备择假设中预测的最小的秩和，R_2 是备择假设中预测的第二小的秩和，依此类推.

$$T_4 = 17 + 2(18) + 3(21) + 4(34.5) + 5(29.5) = 401.5$$

$$T_5 = \frac{401.5 - 8(5)(36)/4}{\left[8(5^3-5)^2/144(4)\right]^{1/2}} = \frac{41.5}{\sqrt{200}} = 2.9345$$

将 T_5 与表 A1 中的值相比较得 $p=0.002$，则在 $\alpha=0.05$ 的水平上拒绝 H_0. 当没有结时，Page 构造的精确表也给出了同样的 p-值.

我们也可以代替 Spearman ρ 使用 Kendall τ(Jonckheere,1954b). Shorak(1967)和 Pirie,Hollander(1972)对相同的假设给出了其他的非参数检验方法. ■

下面的讨论表明 Friedman 检验统计量和其他一些常用的非参数统计量有着密切
381 的关系. 为了简便，下面的讨论仅限于没有结的情况. 对存在结的情况，我们可以进行类似的比较.

与 Kendall 协调系数的关系

Kendall 和 Babington-Smith(1939)以及 Wallis(1939)分别独立地介绍了统计量 W，称为 Kendall 协调系数. 它可以应用于 Friedman 检验统计量应用的相同情况，尽管起初可能想把它作为 b 个区组中“秩一致性”的度量，而不是一个检验统计量. 用与前面相同的符号，Kendall W 定义如下：

$$W = \frac{12}{b^2k(k+1)(k-1)} \sum_{j=1}^{k} \left[R_j - \frac{b(k+1)}{2} \right]^2 \tag{23}$$

如果在 b 个区组中的秩完全一致，那么处理1 在全部 b 个区组中得到相同的秩，处理 2 在全部 b 个区组中也得到相同的秩，依此类推，此时得到 W 的值等于 1.0. 如果秩中有“明显的不和谐”，则 R_j 的值要么相等，要么相互十分接近，且接近于它的均值，因此 W 将等于 0 或非常接近于 0.

比较 Kendall W 和由（5）式导出的 Friedman 检验统计量，我们可以得到如下关系

$$W = \frac{T_1}{b(k-1)} \tag{24}$$

因此，W 只是 Friedman 检验统计量的一个简单的变形，对于任何用 W 作为检验统计量的假设检验都可以用计算 T_1 替代 W. 如果 T_1 超过了它的零分布的 $1-\alpha$ 分位数，则 W 也会超过它自身零分布的 $1-\alpha$ 分位数. *Minitab*, *StatXact* 和 *SPSS* 都有计算 W 的程序.

与 Spearman ρ 的关系

由（5.4.4）式定义的 Spearman ρ，可以在两个区组之间进行计算，例如区组 i 和区组 m，将这两区组作为两个样本来考虑，且将在每个处理之下的两个秩看作是一对相关的秩. 对所有区组对的 Spearman ρ 的平均值和 Friedman 检验统计量有一个直接关系，我们现在来验证.

记 ρ_a 为 Spearman ρ 的平均值，即要对 $b(b-1)$ 个 ρ_{im} 的值求平均，因为即使由对称性有 $\rho_{im}=\rho_{mi}$，也要计数 ρ_{im} 和 ρ_{mi} 的个数. 为了计算 ρ 的平均值，我们将对所有的 i, m 求和，然后减去当 i 和 m 相等时的 ρ_{im}；即我们要减去和自己配对区组的 ρ 值. 共有 b 个情况下的 ρ_{im} 等于 1. 因此，ρ 的平均值表示如下： 382

$$\rho_a = \frac{1}{b(b-1)} \left(\sum_{i=1}^{b} \sum_{m=1}^{b} \rho_{im} - b \right) \tag{25}$$

如果秩之间存在“完全的一致”，在上述描述的意义下，因为每个 ρ_{im} 等于 1，所以随机变量 ρ_a 等于 1. 如果秩之间不一致，ρ_a 将小于 1 或甚至会有负值. 然而，除了只有两个区组（$b=2$）这一特殊情况外，ρ_a 不可能小于 -1.

（25）式和 Spearman ρ 的定义可以组合并简化，以揭示出它与（2）式给出的

Friedman 检验统计量 T_1 的关系：

$$\rho_a = \frac{T_1}{(b-1)(k-1)} - \frac{1}{b-1} \tag{26}$$

因此，平均 Spearman ρ 的分位数可以简单地从 Friedman 检验统计量的分位数得到.

上述的 Friedman 检验统计量和 Spearman ρ 之间的关系说明，Friedman 检验可以用作两样本情形中线性相关的检验，Spearman ρ 已用在这个检验中. 尽管 Friedman 检验统计量的精确分布可以很容易从 Spearman ρ 的分布得到，Spearman ρ 也有一个对小样本可以制表的优点. 两个检验是等价的，因此，在 Pearson r 适用的情况下，相对于通常用 Pearson r 作为检验统计量的参数检验，它们两个的 A. R. E. 都是0.912.

每个试验单元中几个观测情形的推广

如果在每个区组内，每个处理有几个（m）观测，而不是像以前每个试验单元只有一个观测的情形，处理之间没有差异的零假设将由稍微修改的 Friedman 方法来检验. 像前面一样，对每个区组内的观测进行排序，不同的是，秩从 1 到 mk. 如前，秩 R_j 的和定义为赋给所有观测（包括处理 j）的秩和. 在区组 i 中用处理 j 的观测记为 $X_{ij1}, X_{ij2}, \cdots, X_{ijm}$. R_j 的均值变为

$$\begin{aligned} E(R_j) &= \sum_{i=1}^{b}\sum_{n=1}^{m} E[R(X_{ijn})] = \sum_{i=1}^{b}\sum_{n=1}^{m} \frac{mk+1}{2} \\ &= \sum_{i=1}^{b} \frac{m(mk+1)}{2} = \frac{bm(mk+1)}{2} \end{aligned} \tag{27}$$

由定理1.4.5，可求得 R_j 的方差（这里 n 由 m 代替，N 由 mk 代替）如下：

383

$$\begin{aligned} \mathrm{Var}(R_j) &= \sum_{i=1}^{b} \mathrm{Var}\left[\sum_{n=1}^{m} R(X_{ijn})\right] = \sum_{i=1}^{b} \frac{m(mk+1)(mk-m)}{12} \\ &= \frac{bm^2(mk+1)(k-1)}{12} \end{aligned} \tag{28}$$

如果数据中有结，则 $\mathrm{Var}(R_j)$ 给出如下：

$$\mathrm{Var}(R_j) = \frac{m(k-1)}{k(mk-1)}\left[\sum_{\text{所有秩}} R(X_{ijn})^2 - mkb(mk+1)^2/4\right] \tag{29}$$

对所有的 j，它是相同的.

这里用到检验统计量

$$T_6 = \sum_{j=1}^{k} \frac{(k-1)}{k} \frac{[R_j - E(R_j)]^2}{\mathrm{Var}(R_j)} \tag{30}$$

其中，R_j 的均值和方差如上给出. 像以前一样，要用到自由度为 $k-1$ 的 χ^2 分布表. 多重比较用下列不等式

$$|R_j - R_i| > t_{1-\alpha/2}\left\{\frac{2kb(mk-1)\,\mathrm{Var}\,(R_j)}{(k-1)(mbk-k-b+1)}\left[1-\frac{T_6}{b(mk-1)}\right]\right\}^{1/2} \tag{31}$$

其中，$t_{1-\alpha/2}$是自由度为 $mbk-k-b+1$ 的 t 分布的 $1-\alpha/2$ 分位数，它可从表 A21 中得到.

交互作用

据我所知，对交互作用，没有什么好的、精确的非参数方法. 精确的检验或者要求当区组效应被区组内 Friedman 型秩消除时（例如见 Patel 和 Hoel,1973），处理之间没有差异，或者，它们是基于观测条件是“条件分布自由的”. 后面一组检验有过多的计算量，且包含列秩检验，其中，通过从观测中减掉处理均值或中位数来“消除”处理效应，通过在区组内排序或者从观测中减掉区组均值或中位数来消除区组效应. 根据好的功效和应用方便，对这些列秩检验的大样本逼近，尽管不是真正的非参数检验，但也是参数方法中最好的选择. 它们降低了非正态分 384
布带来的功效问题，且当每个单元的观测趋于无穷时，它们是渐近分布自由的. 为了对交互作用的一些最好的列秩检验进行很好的比较，可参见 Mansouri 和 Chang (1995).

如果备择假设指定有处理效应的排序，那么一些双向表秩和检验的参考书目包括 Page(1963), Hollander(1967b)，和 Pirie(1974)；对于多重比较，有 Dunn (1964)以及 McDonald 和 Thompson(1967). Doksum(1967), Puri 和 Sen(1967), Sen (1968b)以及 Lemmer, Stoker 和 Reinach(1968)建议了其他分析方法. Mehra 和 Sarangi(1967)以及 Sen(1967a)研究了渐近效率. 小样本效率则由 Gilbert(1972)研究. 由 Gerig(1969,1975)考虑了多元情形的推广. Koch(1970)讨论了每个单元格有几个观测的裂区变差. Li 和 Schucany(1975)以及 Schucany 和 Beckett(1976)考虑了度量两个集合区组秩之间的协调性. 应用到双向表，且比 Friedman 检验有更高 A. R. E. 的“列秩”方法的完整表达，可以参看 Lehmann(1975). Hora 和 Iman (1988)给出了列秩检验和 Friedman 检验的 A. R. E. 的比较，有兴趣的读者也可以参见 5.12 节.

习题

1. 对某城市的所有 7 所医院做调查，得到出生 12 个月以上婴儿的人数. 把这一时间段分为 4 个季度，去检验假设 4 个季度的出生率是常数.

   调查结果如下：

	出生人类			
医院	冬季	春季	夏季	秋季
A	92	112	94	77
B	9	11	10	12
C	98	109	92	81
D	19	26	19	18
E	21	22	23	24
F	58	71	51	62
G	42	49	44	41

(a) 用 Friedman 检验分析这些数据.

(b) 用 Quade 检验分析这些数据.

(c) 你能说明这两个检验结果中大的差异吗?

2. 随机选择12个学生参加一个学习试验，试验者制作了4组单词，每组包含有20对单词，
385 但是，在这4组中用不同的配对方法. 每个学生拿一组，给他们5分钟去学习，然后检查他或她的记忆单词的能力. 对每个学生将这个方法对所有4组单词重复，组的顺序从一个学生转到下一个. 测验的得分如下（20是满分）.

	学生											
组	1	2	3	4	5	6	7	8	9	10	11	12
1	18	7	13	15	12	11	15	10	14	9	8	10
2	14	6	14	10	11	9	16	8	12	9	6	11
3	16	5	16	12	12	9	10	11	13	9	9	13
4	20	10	17	14	18	16	14	16	15	10	14	16

有些组单词比其他组容易记吗?

(a) 用 Friedman 检验.

(b) 用 Quade 检验.

3. 用 T_1 重新做例5.8.2，并比较 p-值.

4. 用一个测力计度量机动车的一氧化二氮散发速率，司机在15分钟时间内遵循固定的加速度、减速度等规则. 为了看由于不同的司机，度量是否不同，对6个机动车进行重复检验，每个站有3个司机，结果如下.

	司机		
车	1	2	3
1	6.2	6.3	6.0
2	12.6	12.9	12.7
3	10.2	10.6	9.8
4	13.0	13.1	13.0
5	5.6	5.9	5.5
6	8.1	8.1	7.8

某些司机倾向于比其他司机得到更低的速率吗? 是哪个司机?

5. Fox 和 Randall(1970), Hollander 和 Wolfe(1973)所做的一个试验想说明：增加重量会减少物体的前臂颤动频率. 测量了6种物体，每种都有5个不同的重量，测量前臂颤动频率的数据如下. 问这些数据支持这个理论吗?

	重量(1b.)				
物体	0	1.25	2.5	5	7.5
1	3.01	2.85	2.62	2.63	2.58
2	3.47	3.43	3.15	2.83	2.70
3	3.35	3.14	3.02	2.71	2.78
4	3.10	2.86	2.58	2.49	2.36
5	3.41	3.32	3.08	2.96	2.67
6	3.07	3.06	2.85	2.50	2.43

386

6. 对下面的数据用 *StatXact* 来解释 Page 检验. 人们期望：降低应用于棉植物的碳酸钾水平能倾向于增加光纤的强度. 5 个不同剂量的水平应用于 3 个不同区组. *StatXact* 给出 $p = 0.0025$. 这和你的分析怎么比较？

	碳酸钾的水平(1b/英亩)				
区组	144	108	72	54	36
1	7.46	7.17	7.76	8.14	7.63
2	7.68	7.57	7.73	8.15	8.00
3	7.21	7.80	7.74	7.87	7.93

思考题

1. 对 $k=2$，证明统计量 T_3 是由（5.7.5）式给出的 Wilcoxon 符号秩检验统计量的函数，因此，两个检验是等价的（提示：首先证明 Q_i 等于 R_i 的绝对值）.
2. 对 $k=2$，证明 Friedman 检验等价于双边符号检验（用大样本逼近）.

5.9 平衡的不完全区组设计

在 5.8 节开头所描述的随机的完全区组设计中，每一个区组都应用每一个处理. 但是在现实中，往往很难做到在每一个区组中应用所有的处理，特别是当处理的数量较大，而区组的大小有限时. 例如，如果需要品尝 20 种食品，而每个品尝者（区组）往往发现实际上很难对这 20 种食品进行精确的排序. 但是如果使用 4 倍的人（或者每个人用 4 次），而每一个人只需品尝 5 种食品，那么评判起来就会更加容易和精确. 像这样的试验设计，每一个区组中没有用所有的处理就叫做不完全的区组设计. 而且，如果设计满足下述条件，我们就称它是*平衡的*（balanced）不完全区组设计：(1) 每一个区组包括 k 个试验单元，(2) 每一个处理出现在 r 个组中，(3) 每一个处理与其他处理出现的次数相同.

Durbin(1951)提出了一个秩检验可以用于检验平衡的不完全区组设计中的零假设，即不同的处理间没有显著差异. 我们已有一些参数方法来分析用平衡的不完全区组设计所得的数据，这些方法都是基于正态分布的假设基础上的，我们这里不做更多的解释. 如果正态分布的假设无法得到满足，而且我们想要简单的分析方法， 387
或者我们得到的观测只有秩的话，那么 Durbin 检验往往比参数方法更受欢迎. 如果处理数和每个区组中试验单元数相等，那么 Durbin 检验可以简化为 Friedman 检验.

如果所给的第 3 个条件不是完全满足，Durbin 检验在大多数情况下仍然有效.

▶Durbin 检验

数据　我们这里使用如下符号.

t = 所考查的处理数.

k = 每一个区组中试验单元数（$k<t$）.

b = 区组总数.

r = 每一个处理出现的次数（$r<b$）.

λ = 区组中同时出现第 i 个处理和第 j 个处理的区组数.（对于任意一对处理，λ 的值相等.）

把数据配置在上面定义的平衡的不完全区组设计中，并令 X_{ij} 代表区组 i 中处理 j 的结果（若处理 j 出现在区组 i 中）.

设每个区组中只有 k 个观测结果，且给每一个区组中的 X_{ij} 赋以秩，其中秩 1 赋给区组 i 中最小的观测值，秩 2 赋给区组 i 中第二小的观测值，依此类推直到 k，表示区组 i 中最大观测值的秩. 如果 X_{ij} 存在，则记 $R(X_{ij})$ 为 X_{ij} 的秩.

计算第 j 个处理下的 r 个观测值的秩和，并记这个和为 R_j，则 R_j 可写成

$$R_j = \sum_{i=1}^{b} R(X_{ij}) \tag{1}$$

其中，在处理 j 下，只有 r 个 $R(X_{ij})$ 的观测值存在，因此 R_j 只是 r 个项的秩和.

如果观测值不是数值，但可以根据某些感兴趣的原则，在区组内对对象进行排序，赋予每一个观测以相应的秩，并如上计算出 $R_j, j=1,2,\cdots,t$.

如果由于某些观测值相等导致有几种不同的赋秩方法，则我们推荐运用赋平均秩到每个有结观测的方法. 这个方法可能会改变统计量的零假设，但是如果结的数
388 量不多的话，可以忽略其影响.

假定条件

1. 区组相互独立.
2. 每一个区组中观测值具有次序度量尺度，结不会导致出问题.

检验统计量　Durbin(1951)建议使用如下检验统计量

$$T_1 = \frac{12(t-1)}{rt(k-1)(k+1)} \sum_{j=1}^{t} \left(R_j - \frac{r(k+1)}{2}\right)^2 \tag{2}$$

如果区组中存在结，则使用赋平均秩的方法，并进行调整. 记 A 为秩与平均秩的平方和，

$$A = \sum_{i=1}^{b} \sum_{j=1}^{t} [R(X_{ij})]^2 \tag{3}$$

同时计算“校正因子”C 如下

$$C = \frac{bk(k+1)^2}{4} \tag{4}$$

由于存在结，调整后的统计量 T_1 变成：

$$T_1 = \frac{(t-1)\sum_{j=1}^{t}\left(R_j - \frac{r(k+1)}{2}\right)^2}{A-C} = \frac{(t-1)\left[\sum_{j=1}^{t} R_j^2 - rC\right]}{A-C} \tag{5}$$

另外一个等价的方法是在秩与平均秩上使用通常的方差分析法. 这样就得出了下面的统计量 T_2，它仅是 T_1 的一个函数. 最近的研究表明，T_2 的近似分位数比 T_1 的更精确一些，因此人们更愿意使用统计量 T_2.

$$T_2 = \frac{T_1/(t-1)}{(b(k-1)-T_1)/(bk-b-t+1)} \tag{6}$$

总的来说，用 T_2 作为检验统计量，首先要计算统计量 T_1，如果没有结，就使用（2）式；如果有结，则使用（3），（4），（5）式.

零分布 由于 T_1（或 T_2）的精确分布很难求出，我们往往使用它们的逼近分布. T_1 的逼近分布是一个自由度为 $t-1$ 的 χ^2 分布，这个逼近分布趋向于保守. T_2 的逼近分布是一个自由度为 $k_1 = t-1, k_2 = bk-b-t+1$ 的 F 分布（见表 A22），这个逼近分布 389
倾向于给出一个膨胀的 α 值，但是比 T_1 更加接近所需要的值.

假设

H_0:每一个区组中，所有随机变量的赋秩都是等可能的（即处理有相同的效应）

H_1:至少有一个处理倾向于产生比至少一个其他处理有较大的观测值

如果 T_2 大于由表 A22 查出的 F 分布的 $1-\alpha$ 分位数，其中自由度为 $k_1 = t-1$，$k_2 = bk-b-t+1$，则我们以近似水平 α 拒绝 H_0. 我们也可以用表 A22 得到近似的p-值.

多重比较 如果检验中拒绝了零假设，则可以使用处理对间的多重比较，过程如下. 考虑处理 i 和处理 j，如果它们的秩和 R_i 与 R_j 满足如下不等式：

$$|R_i - R_j| > t_{1-\alpha/2}\left[\frac{(A-C)2r}{bk-b-t+1}\left(1 - \frac{T_1}{b(k-1)}\right)\right]^{1/2} \tag{7}$$

则认为处理 i 和处理 j 有不同的均值，其中，A, C 和 T_1 由（3），（4），（5）式给出，$t_{1-\alpha/2}$是自由度为 $bk-b-t+1$ 的 t 分布的 $1-\alpha$ 分位数，它可以从表 A21 得到. 如果没有结，则（7）式可以化简为

$$|R_i - R_j| > t_{1-\alpha/2}\left[\frac{rk(k+1)}{6(bk-b-t+1)}(b(k-1)-T_1)\right]^{1/2} \tag{8}$$

计算机辅助 对平衡的不完全区组设计的分析，可以首先在每一区组中将数据转化为秩，然后应用软件中的相应计算机程序进行计算，如 *SAS* 中用于秩上的参数平衡的不完全区组设计程序，或是 *Minitab*, *SPSS*, *SAS* 中用于秩上的广义线性模型的计算程序. ◀

例 5.9.1

假设冰淇淋生产厂商希望测试 7 种不同口味的冰淇淋受人们喜欢的程度. 他让每一个受测试者品尝 3 种不同的冰淇淋，并用 1,2,3 将其赋秩，秩 1 代表最喜欢的种类. 为了设计一个试验，使得每一种冰淇淋被品尝的次数相同，我们使用了 Youden 方阵（Federer,1963）来编排. 7 个人分别品尝了 3 种不同的冰淇淋，得到的排序结果
390 如下：

				种类			
人	1	2	3	4	5	6	7
1	2	3		1			
2		3	1		2		
3			2	1		3	
4				1	2		3
5	3				1	2	
6		3				1	2
7	3		1				2
$R_j =$	8	9	4	3	5	6	7

在这个试验中，

$t=7$ = 冰淇淋种类的总数

$k=3$ = 一次所比较的冰淇淋种类的数量

$b=7$ = 品尝者（区组）的总数

$r=3$ = 每一种冰淇淋被品尝的次数

$\lambda=1$ = 每种冰淇淋与其他种类冰淇淋比较的次数

因此，这是一个平衡的不完全区组设计，我们使用 Durbin 的方法检验零假设：7 种冰淇淋受喜欢的程度相当.

因此在得到近似水平 $\alpha=0.05$ 下，检验的临界域对应着 T_2 值大于 3.58，其中 3.58 为自由度分别为 $k_1=t-1=6, k_2=bk-b-t+1=8$ 的 F 分布的 0.95 分位数（由表 A22 获得）.

由于没有结存在，我们首先根据（2）式，求得统计量 T_1：

$$T_1=\frac{12(t-1)}{rt(k-1)(k+1)}\sum_{j=1}^{t}\left[R_j-\frac{r(k+1)}{2}\right]^2$$

$$=\frac{(12)(6)}{(3)(7)(2)(4)}[(8-6)^2+(9-6)^2+\cdots+(7-6)^2]$$

$$=12$$

然后由（6）式，得到统计量 T_2

$$T_2=\frac{T_1/(t-1)}{(b(k-1)-T_1)/(bk-b-t+1)}$$

$$=\frac{12/6}{(14-12)/8}=8$$

由于 T_2 属于临界域，故拒绝零假设. 由表 A22 可得出 p-值小于 0.01 .

利用（8）式得到的多重比较表明，第 4 种冰淇淋比第 3 种和第 7 种更受喜欢，而第 3 种冰淇淋比第 7 种更受喜欢. 391

注意在本例中，7 位参与者品尝冰淇淋具有很好的一致性，因此我们可以比较容易地找到精确的 p-值. 每一位品尝到第 4 种口味时都会表示他更加喜欢它，每一位品尝到第 3 种时会表示除了第 4 种，他会更喜欢这一种，而每一位品尝到第 5 种时会表示除了第 3 种和第 4 种，他会更喜欢这一种，依次类推. 人们由此达成了很好的一致性，即冰淇淋的受喜欢程度由强至弱依次为 4,3,5,6,7,1,2. 也就是说，如果冰淇淋口味种类没有差异的零假设成立，则对每一个人来说，对于他所品尝的 3 种冰淇淋都会出现 $3!=6$ 种等可能的排序方式，那么 7 个人共有 $6^7=279936$ 种等可能的排序组合. 而其中只有一种可能的排序出现了，即第 4 种最受喜欢，第 3 种其次，等等. 但是，还有其他达成一致方式的可能性发生，如在 7 种中选出最受喜欢的，然后在其余 6 种中选出次受喜欢的，等等，共有 7! 种可能达成一致的方式. 所以本例中达成一致的概率为

$$P(T_2 \geqslant 8) = P(T_2 = 8) = \frac{7!}{6^7} = 0.018$$

即我们得到的精确的 p-值 0.018，它略大于由表 A22 得出的近似 p-值（小于 0.01）. 但是它比由 T_1 的逼近分布（即自由度为 6 的 χ^2 分布）所得到的近似 p-值（大于 0.05）准确得多. ■

□**理论** Durbin 检验和 Friedman 检验的理论发展十分相似，因为在处理间无差异的零假设下，区组内 k 个秩的排列方式是等可能出现的，因此可以求得 Durbin 检验统计量的精确分布. 在每一个区组中有 $k!$ 种等可能的秩排列方式，共有 b 个组. 因此从 b 个区组的总体来看，因为共有 $(k!)^b$ 种可能不同秩的排列方式，且每一种秩的排列方式都是等可能的，出现的概率为 $1/(k!)^b$. 如同前一节中的 Friedman 检验一样，对每一种排列方式，计算 Durbin 检验统计量，从而确定它的分布函数.

在许多情况下，求 Durbin 检验统计量 T_1 的精确的分布不太实际，所以如果每一种处理的重复数 r 比较大，它的分布往往由自由度为 $t-1$ 的 χ^2 分布逼近. 我们对这个逼近进行调整如下：

如果每一种处理的重复数 r 比较大，那么根据中心极限定理，第 j 个处理下的秩和 R_j 可以逼近正态分布. 因此随机变量

$$\frac{R_j - E(R_j)}{\sqrt{\operatorname{Var}(R_j)}}$$

392

近似服从标准正态分布. 正如前一节所述，如果 R_j 是独立的，那么统计量

$$T' = \sum_{j=1}^{t} \frac{[R_j - E(R_j)]^2}{\operatorname{Var}(R_j)} \tag{9}$$

可以近似看作是 t 个独立的近似服从 χ^2 分布（自由度为1）的随机变量的和，则 T' 近似服从一个自由度为 t 的 χ^2 分布. 但是，这里 R_j 并不独立，它们的和为一定值：

$$\sum_{j=1}^{t} R_j = \frac{bk(k+1)}{2} \tag{10}$$

因此，如果知道了 $t-1$ 个 R_j，我们就可以求得余下的一个 R_j. Durbin(1951)证明了用 $(t-1)/t$ 乘以 T' 得到一个统计量，它近似服从自由度为 $t-1$ 的 χ^2 分布，其形式表示为

$$T_1 = \frac{t-1}{t}T' = \frac{t-1}{t}\sum_{j=1}^{t}\frac{[R_j - E(R_j)]^2}{\operatorname{Var}(R_j)} \tag{11}$$

为了把（11）式转换为常见的由（2）式给出的形式，我们只需要找到 R_j 的均值和方差.

秩和 R_j 为独立随机变量 $R(X_{ij})$ 的和，

$$R_j = \sum_{i=1}^{b} R(X_{ij}) \tag{12}$$

每一个 $R(X_{ij})$ 只要它存在，就是从整数1到 k 中的一个随机选取. 因此，根据定理1.4.5，可以得到 $R(X_{ij})$ 的均值和方差如下：

$$E[R(X_{ij})] = \frac{k+1}{2} \tag{13}$$

和

$$\operatorname{Var}[R(X_{ij})] = \frac{(k+1)(k-1)}{12} \tag{14}$$

那么，可以容易地求得 R_j 的均值和方差为：

$$E(R_j) = \sum_{i=1}^{b} E[R(X_{ij})] = \frac{r(k+1)}{2} \tag{15}$$

和

393

$$\operatorname{Var}(R_j) = \sum_{i=1}^{b}\operatorname{Var}[R(X_{ij})] = \frac{r(k+1)(k-1)}{12} \tag{16}$$

将 R_j 的均值和方差代入（11）式，得到，

$$\begin{aligned} T_1 &= \frac{t-1}{t}\sum_{j=1}^{t}\frac{[R_j - r(k+1)/2]^2}{r(k+1)(k-1)/12} \\ &= \frac{12(t-1)}{rt(k+1)(k-1)}\sum_{j=1}^{t}\left[R_j - \frac{r(k+1)}{2}\right]^2 \end{aligned} \tag{17}$$

这与在关于Durbin检验的解释中所给的具有同样形式.

χ^2 分布逼近是基于每一种处理的重复次数 r 相当大这一假设的，但在实际情况中，重复的次数 r 有时会比较小，如3或2，此时如果用 χ^2 分布逼近，则所说的 α 水平可能并不是很精确. 而如果应用基于秩的方差分析统计量 F（即（6）式给出的统计量 T_2），则所说的 α 水平更接近真正的 α 水平.

(Benard 和 van Elteren, 1953) 已把 Durbin 检验推广到一个试验单元里有若干个观测的情形. Noether(1967a)也讨论了 Durbin 检验以及它的推广, 并证明了 Durbin 检验相对于它的参数检验情形的 A. R. E. 与 Friedman 检验相对于它的参数检验情形的 A. R. E. 相同. Puri 和 Sen(1969b)讨论了成对比较 ($k=2$) 的情况. □

习题

1. 通过试验来检验 7 种类型轮胎的耐用性. 一般认为最好的试验方法是观测在实际驾驶的情况中轮胎的表现. 但是, 一次只有 4 个轮胎可以同时进行比较, 因为一辆汽车只有 4 个轮子可以进行试验. 因此, 试验设计采用平衡的不完全区组设计. 对 7 位司机, 每一位随机地选取所驾驶汽车的 4 个轮胎类型, 并在试验中有规律地轮换. 轮胎在必要的时候进行更换, 根据更换的顺序赋秩给原始的轮胎.

	轮胎类型						
司机	1	2	3	4	5	6	7
1			3		1	4	2
2	1			3		4	2
3	2	1			3		4
4	1	2	4			3	
5		1	4	3			2
6	2		4	1	3		
7		1		2	3	4	

394

结果是否显示出耐用性有显著差异? (首先检查试验是否服从一个平衡的不完全区组设计). 如果耐用性有显著性差异, 那么, 用多重比较方法确定哪一个轮胎类型优于其他的轮胎类型.

2. 为了对食肉动物进行控制, 设计一个试验用于确定 5 种气味哪一种对丛林狼更有吸引力. 试验者已经发现在同一时间至少存在 3 种气味才会迷惑丛林狼, 并且产生不一致的结果. 因此, 一次把 3 种气味放在一个大的开阔地的不同地方, 一次放出一只丛林狼到开阔的地方, 记录它在每一种气味的地方停留的时间 (以秒为单位).

   根据平衡的不完全区组设计, 轮换 3 种气味, 得到如下结果.

	气味				
丛林狼	1	2	3	4	5
1	12	23		14	
2		17	2		2
3	16		1	6	
4		42		10	0
5	8		6		1
6	22	31			0
7	28	16	4		
8	15			7	4
9		67	5	18	
10			6	16	1

这是一个平衡的不完全区组设计吗？不同的气味之间有显著差异吗？如果有，哪一种气味优于其他的气味？

3. 将一个金融班的学生分成5组，每一个组完成一个课题，并且在班中其他组前做一个报告，做完所有的报告以后，每一个组给其他的组评分，从最好的（10分）到最差的（0分）打分. 这里是他们的评分.

	被评组				
评分组	1	2	3	4	5
1		6.7	9.1	8.6	9.2
2	7.6		9.0	8.1	9.3
3	8.6	8.3		8.9	9.4
4	8.9	8.5	8.8		9.6
5	9.1	9.3	9.6	9.4	

每一个组得到的评分是否有显著区别？如果有，哪一个组得到的评分显著高于其他的组？

395

思考题

1. 证明：$kb = rt$（提示：以两种不同方法计数观测值数）.
2. 证明：$\lambda = r(k-1)/(t-1)$（提示：首先注意任意一个制定的处理发生在 r 区组中，然后计算那些处理没有出现在这 r 个区组里的单元数，并用两种不同的方法计数）.

5.10　A. R. E. 不低于1的检验

本节所描述的检验具有一个共同的性质，当它们与通常的参数检验比较时，只要参数检验是合适的，它们的 A. R. E. 都等于1. 如果参数检验中的正态分布假设不满足，那么在某些常见的条件下，A. R. E. 往往大于1，甚至趋近于正无穷. 如果用渐近相对效率（A. R. E.）来衡量检验，则本节中的检验总是要优于一般的参数检验，如 t 检验，F 检验，这个结论听起来似乎相当强，但这是事实. 记住，A. R. E. 只是许多衡量检验方法中的一个，尽管它也是比较检验的方法中最被广泛接受的一个. 相对效率（无渐近）也是一个比较方法，当样本容量有限时，在同等条件下，它比较两个检验如果具有相同的功效时所需要的样本大小. 根据相对效率，本章中的检验视具体情况情况可能优于或劣于它们的参数检验情形. 由于我们很难考虑到所有的情况，所以我们通常使用 A. R. E. 来比较检验.

与本章前面几节不同，本节中我们将不介绍新的试验情况. 我们已经介绍了基于秩的非参数方法，以便解决5.2节的单向表，5.4节的相关性，以及5.8节的随机化的完全区组设计，这些方法都是被广泛接受的，合理有效而且不难操作. 相比较而言，本节中的方法一样有效，只是操作起来略有难度. 本节中的假设实际上等同于前面检验的基本假设. 事实上，它们本质上还是秩检验，只是略有修饰使得它们

的A.R.E. 更好. 使用者可以自己决定是否用这些和前面介绍的检验，没有严格的统计基础说明一定要使用某种检验.

我们介绍的第一种检验是基于 van de Waerden(1952,1953)建议的一个简单想法. 即在所有计算中，用另外一些数据替换基本数据的秩，如近似服从正态分布的数据，特别地，我们可利用标准正态分布的 $k/(N+1)$ 分位数，其中，$k=1,2,\cdots,N$，N 代表样本容量，这些分位数有时称作正态得分，可从表A1中获得. 比如，有5个观测值的随机样本，由小到大依次为7.3,7.7,9.2,12.0和26.4. 注意较小的3个观测值差别不大，第4个观测值较前3个明显偏大，而最大的观测值是所有观测值的至少2倍. 用它们的秩1,2,3,4,5替代它们的观测值，也就是把非对称的原始数据转化为对称的、近似等分的、像“均匀分布”的数据. 在本章的前几节中，我们解释了对原始数据通常的同类分析如何也能转化成对于秩的分析. 现在，根据 van der Waerden 的建议，我们将秩转化为正态得分，即用由表A1中得到的正态分布的 $k/(N+1)$ 分位 396
数取代秩 k，因此秩1将转化为 $z_{1/6}=z_{0.167}=-0.9661$，秩2将转化为 $z_{2/6}=z_{0.333}=-0.4316$，依此类推. 然后我们不再分析这些秩，而是分析得到的正态得分：$-0.9661,-0.4316,0.0000,0.4316$ 和 0.9661. 一般来说，这些得分以零为中心对称分布，并将与“完美的正态样本”具有相似的散布区域（“完美的正态样本”当然是没有的）. 正态得分检验是一个非参数检验，它的渐近效率与总体是正态分布前提下的参数检验相同；而在总体是非正态分布时，它有较大的渐近效率.

下面我们将通过分析显示，看怎样用正态得分来作为5.2节中检验 k 个总体相同的 Kruskal-Wallis 检验的一个调整，两样本问题是5.1节中的有关 Mann-Whitney 检验的特殊情形.

▶几个独立样本的 van der Waerden（正态得分）检验

数据 数据由 k 个随机样本组成，每一个样本可能具有不等的样本容量. 记第 i 个容量为 n_i 的样本为：$X_{i1},X_{i2},\cdots,X_{in_i}$. 令 N 表示样本观测总数. 如 Kruskal-Wallis 检验所述，给 N 个观测值从秩1到秩 N 赋秩，当存在结时，使用平均秩，并记 X_{ij} 的秩为 $R(X_{ij})$.

变换每个秩 R 为标准正态分布的 $R/(N+1)$ 分位数（见表A1）. 为了简便，称这些分位数为“正态得分”，记作 A_{ij}.

$$A_{ij}=z_{R(X_{ij})/(N+1)}=\text{从表A1得到的第}\frac{R(X_{ij})}{N+1}\text{个分位数} \tag{1}$$

为了便于得到正态得分，我们计算出 $R(X_{ij})/(N+1)$ 后只保留3位小数，然后通过表A1求值，则 k 个样本中每个样本的平均得分为：

397

$$\overline{A}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} A_{ij} \qquad i = 1, 2, \ldots, k \tag{2}$$

方差为

$$S^2 = \frac{1}{N-1}\sum_{\substack{所有\\得分}} A_{ij}^2 \tag{3}$$

注意，如果没有结或有许多结，但本质上结为零时，则总体均值等于零. 因此在计算方差时，总体均值可以省略.

假定条件 这里的假定条件与 Kruskal-Wallis 检验的假定条件相同.

检验统计量 检验统计量 T_1 定义为

$$T_1 = \frac{1}{S^2}\sum_{i=1}^{k} n_i(\overline{A}_i)^2 \tag{4}$$

其中，$\bar{A}_i$ 和 S^2 分别如（2），（3）式所示.

零分布 正如在分析秩 $R(X_{ij})$ 的所有置换后，得到 Kruskal-Wallis 检验统计量的精确分布一样，T_1 的精确分布在分析得分 A_{ij} 的所有置换后也可以得到. 但是，这样的分析在许多情况下是很困难的，所以我们经常用自由度为 $k-1$ 的 χ^2 分布作为逼近分布，而这样的逼近通常是很好的.

假设 如同 Kruskal-Wallis 检验一样，我们有：

H_0: 所有 k 个总体的分布函数相同

H_1: 至少有一个总体倾向于比至少一个其他分布产生较大的观测值

如果通过查表 A2 得到统计量 T_1 超过 χ^2 分布（自由度为 $k-1$）的 $1-\alpha$ 分位数，那么以水平 α 拒绝零假设. 注意这只是一个近似，但是在实际应用中，它还是足够好的. 我们还可以通过比较 T_1 和表 A2 中的分位数得到 p-值.

多重比较 如果拒绝了零假设，那么我们可以得出结论，如果下列不等式成立

$$|\overline{A}_i - \overline{A}_j| > t_{1-\alpha/2}\left(S^2\frac{N-1-T_1}{N-k}\right)^{\frac{1}{2}}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)^{\frac{1}{2}} \tag{5}$$

则总体 i 和 j 不相同，其中，$t_{1-\alpha/2}$ 为自由度为 $N-k$ 的 t 分布的 $1-\alpha/2$ 分位数，它可以由表 A21 得到，其他项如前所定义. 对所有的 i 和 j 的组合，这个方法可以重复进

398 行，在多重比较中使用的 α 与前面所用的 α 相同.

计算机辅助 *StatXact* 中含有正态得分检验的程序，可以求出精确的 p-值. ——◀

例 5. 10. 1

再来考虑 5. 2 节中用于说明 Kruskal – Wallis 检验的例题，在此我们还将用到 4. 3 节中的中位数检验来进行这些方法的比较.

下表列出了用 4 种不同方法种植玉米谷物所得到的观测值和它们的秩.

方法1			方法2			方法3			方法4		
观测	秩	正态得分	观测	秩	正态得分	观测	秩	正态得分	观测	秩	正态得分
83	11	−0.4845	91	23	0.4043	101	34	1.8957	78	2	−1.5805
91	23	0.4043	90	19.5	0.1434	100	33	1.5805	82	9	−0.6526
94	28.5	0.8927	81	6.5	−0.8927	91	23	0.4043	81	6.5	−0.8927
89	17	−0.0351	83	11	−0.4845	93	27	0.7421	77	1	−1.8957
89	17	−0.0351	84	13.5	−0.2898	96	31.5	1.2816	79	3	−1.3658
96	31.5	1.2816	83	11	−0.4845	95	30	1.0669	81	6.5	−0.8927
91	23	0.4043	88	15	−0.1789	94	28.5	0.8927	80	4	−1.2055
92	26	0.6526	91	23	0.4043				81	6.5	−0.8927
90	19.5	0.1434	89	17	−0.0351						
			84	13.5	−0.2898						
平均得分 $\overline{A}_i$:		0.3582			−0.1703			1.1234			−1.1723

用下面的方式将秩转化为正态得分. 总样本容量为 $N=34$, 因此, 用每一个秩除以 $N+1=35$, 结果保留到3位小数. 例如, 第一个观测值的秩为11, 则11/35等于0.314, 查表A1, 我们得到0.314分位数的对应值为 −0.4845.

每一种种植方法的平均得分如上表所示. 由 (3) 式计算方差, 即34个正态得分的平方和除以33, 结果是 $S^2=0.8447$. 一般来讲, S^2 总是略小于1.0. 我们得到的 T_1 的观测值为25.1840, 它远远大于由表A2查到的自由度为 $k-1=3$ 的 χ^2 分布的0.95分位数, 即7.815, 因此, 拒绝零假设. 当用中位数检验 (例4.3.1, 其中统计量 $T=17.6$) 和 Kruskal-Wallis 检验 (例5.2.1, 其中统计量 $T=25.46$) 时, p-值小于0.001.

在进行多重比较的过程中, 我们应用自由度为 $30(=34-4)$ 的 t 分布的0.975分位数, 由表A21查得为2.042. 计算结果如下. 399

	$\lvert\overline{A}_i-\overline{A}_j\rvert$	$t_{0.975}\left(S^2\frac{N-1-T_1}{N-k}\right)^{\frac{1}{2}}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)^{\frac{1}{2}}$
$i=1, j=2$	0.5286	0.4401
$i=1, j=3$	0.7652	0.4828
$i=1, j=4$	1.5305	0.4655
$i=2, j=3$	1.2937	0.4721
$i=2, j=4$	1.0020	0.4544
$i=3, j=4$	2.2957	0.4958

在每一种情况下, 平均得分都足够大, 因此我们得出结论: 任意两个总体都是不相同的. 注意, 在5.2节中, 用 Kruskal-Wallis 检验也得到了相同的结论. 这两种检验的结论经常一致, 但并不总是一致. 为了避免含糊不清的情况出现, 我们通常只用两种检验的其中一种, 而不同时使用. ■

现在应该明确的是, 使用正态得分和秩的方式是相同的: 用数字取代原始数据. 对于正态得分的分析与对于秩的分析是相似的. 我们可以给出精确的表, 但是我们

在此并不给出，而是一概使用大样本逼近方法，无论实际样本是大还是小.

Van Eden(1963)提到了一样本问题的 van der Waerden 类型的检验，它类似于 Wilcoxon 符号秩检验. 用 R_i 表示5.7节中在（5.7.3）式前面定义的符号秩. 如果我们不使用符号秩 R_i，而是使用正态分布的第 $\frac{1}{2}[1+R_i/(n+1)]$ 分位数（见表 A1），注意，n 是数据中非零差异数，则我们称之为符号正态得分，记为 A_i. 注意，A_i 与 R_i 有相同的符号. 因此我们可以用检验统计量

$$T_2=\frac{\sum_{i=1}^{n}A_i}{\sqrt{\sum_{i=1}^{n}A_i^2}} \tag{6}$$

与标准正态分布的分位数相比较，得到 Wilcoxon 符号秩检验的一个近似检验，两个检验具有相同的假设和假定条件，而精确的 p-值可以用 *StatXact* 求得.

例 5.10.2

为了进行比较，我们使用例5.7.1中的数据来检验

H_0:双胞胎中先出生的一个不比后出生的更具有进取性

400 H_1:双胞胎中先出生的一个比后出生的更具有进取性

数据如下：

双胞胎集	第一个出生 X_i	第二个出生 Y_i	差 D_i	$\|D_i\|$的秩	符号秩 R_i	符号正态得分 A_i
1	86	88	+2	3	3	0.3186
2	71	77	+6	7	7	0.8134
3	77	76	−1	1.5	−1.5	−0.1560
4	68	64	−4	4	−4	−0.4316
5	91	96	+5	5.5	5.5	0.6098
6	72	72	0	—	—	—
7	77	65	−12	10	−10	−1.3852
8	91	90	−1	1.5	−1.5	−0.1560
9	70	65	−5	5.5	−5.5	−0.6098
10	71	80	+9	9	9	1.1503
11	88	81	−7	8	−8	−0.9661
12	87	72	−15	11	−11	−1.7279

由于（6）式定义的检验统计量 T_2 等于

$$T_2=\frac{\sum_{i=1}^{n}A_i}{\sqrt{\sum_{i=1}^{n}A_i^2}}=\frac{-2.5405}{\sqrt{8.9027}}=-0.8514 \tag{7}$$

它对应于单边 p-值为0.197（见表 A1），这与 Wilcoxon 符号秩检验的结论一致，因为 Wilcoxon 符号秩检验的 $T=-0.7565$，p-值 $=0.238$. ■

Klotz(1962)介绍了应用正态得分进行两个样本的等方差检验，这个检验开始时像两样本的 van der Waerden 检验，但是，后面在统计量

$$T_3 = \frac{\sum_{i=1}^{n} A_i^2 - \frac{n}{N}\sum_{i=1}^{N} A_i^2}{\left\{\frac{nm}{N(N-1)}\left[\sum_{i=1}^{N} A_i^4 - \frac{1}{N}\left(\sum_{i=1}^{N} A_i^2\right)^2\right]\right\}^{\frac{1}{2}}} \tag{8}$$

中所应用的是正态得分的平方而不是正态得分，其中 A_i 表示正态得分，样本容量分别为 m 和 n，$N = m + n$ 则表示联合样本容量. 将 T_3 与表 A1 中的正态分布分位数进行比较. 如果两组样本来自于均值不同的两个总体，则我们在最初赋秩之前，首先应该分别减去它们各自的均值（如果已知）或样本均值. 在 *StatXact* 中可以编程计算 Klotz 检验的精确 p-值. 401

例 5.10.3

考虑例 5.3.1，我们将分析该例的细节，并与平方秩检验进行比较. 检测一台新机器看是否比现有的机器更加稳定，则检验的零假设为：

H_0:新机器和现有机器的变化量相同

其单边备择检验为：

H_1:新机器有更小的方差

由于不知道总体均值，所以首先用数据减去其样本均值来调整数据. 即产生一个类似于平方秩检验的近似检验.

X_i	$X_i - \overline{X}$	秩	$\frac{R_i}{N+1}$	正态得分 A_i	A_i^2
10.8	.06	8	.615	0.2924	.0855
11.1	.36	11	.846	1.0194	1.0392
10.4	−.34	2	.154	−1.0194	1.0392
10.1	−.64	1	.077	−1.4255	2.0321
11.3	.56	12	.923	1.4255	2.0321
Y_i	$Y_i - \overline{Y}$				
10.8	.01	6	.462	−0.0954	.0091
10.5	−.29	3	.231	−0.7356	.5411
11.0	.21	10	.769	0.7356	.5411
10.9	.11	9	.692	0.5015	.2515
10.8	.01	6	.462	−0.0954	.0091
10.7	−.09	4	.308	−0.5015	.2515
10.8	.01	6	.462	−0.0954	.0091

在 Klotz 检验中，一个基本的衡量变化量的量是所有 A_i^2 的和，从第一个样本（从现有的机器）开始计算. 为了检查它的显著性水平，我们先减去它的均值，然后除以它的标准差，在零假设的条件下，得到

$$T_3 = \frac{6.2280 - 3.2669}{1.2629} = 2.3447 \tag{9}$$

（见（8）式），将 T_3 与表 A1 中的数值进行比较，我们得到单边 p-值约为 0.01，与平方秩检验得到的结论相似. ■

402 为了将正态得分的概念用于回归和相关性分析中，首先用正态得分代替 X_i 的秩，然后同样用正态得分代替 Y_i 的秩. 如果没有出现结，X 变量和 Y 变量的正态得分集合应该相同，就像对每一个变量应用相同的秩 1 到秩 n 的集合. 应用正态得分计算 Pearson 乘积矩相关系数（见（5.4.2）式）. 如果没有结，因为平均得分等于零，则相关系数可以化简为

$$\rho = \frac{\sum_{i=1}^{n} A_i B_i}{\sum_{i=1}^{n} A_i^2} \tag{10}$$

（A_i 和 B_i 分别为赋给 X_i 和 Y_i 的正态得分）.（10）式有时也可以用于存在结的情况，除非当结很多时，此时最安全的方法是转而应用（5.4.2）式和实际所用的正态得分. 5.4 节或 5.6 节中描述的方法可能用来处理这样的得分，不过我们在此不做详细讨论.

对于双向表，回忆 5.8 节中的 Friedman 检验，在每一个区组中对观测值赋秩. 这里我们以通常的方式用正态得分代替这些秩. 令 A_{ij} 记为赋给区组 i 中对应于处理 j 的变量 X_{ij} 的正态得分，A_j 记为对应于处理 j 的正态得分和，类似于 Friedman 检验中的 $R(X_{ij})$ 和 R_j. 此时，检验统计量为

$$T_4 = \frac{k-1}{S^2}\left(\sum_{j=1}^{k} A_j^2\right) \tag{11}$$

其中，

$$S^2 = \sum_{\substack{\text{所有}\\\text{得分}}} A_{ij}^2 \tag{12}$$

将统计量和自由度为 $k-1$ 的 χ^2 分布的分位数相比较（见表 A2），过程与 Friedman 检验中的统计量 T_1（见（5）式）相仿. 其他的细节也与讨论 Friedman 检验相同，只不过对于这个新的检验来说，χ^2 分布的近似程度已经非常好了，因此也就无须使用 F 分布. 多重比较的分析与 5.8 节中所描述的分析相同，只不过在此我们用 T_4 和 S^2 代替了（5.8.8）式中的 T_1 和 $A_1 - C_1$.

到目前为止，如何应用正态得分代替秩的方法应该比较清楚了. 相对于参数检验，结果是正态得分检验有略高的 A. R. E.. 相对于前面几节中介绍的秩检验，视具体情况而定，它的 A. R. E. 可能大于 1 或小于 1. 某些其他的得分检验也可以取代正态得分，并可得到与正态得分检验相同的 A. R. E.，其中两种类型的得分称为“随机
403 正态离差”或“期望正态得分”. 我们下面对它们进行简单的介绍.

随机正态离差

对于来自于任一分布的随机样本 $X_1,\cdots,X_n$，我们用一组看似来自正态分布的数来代替它，具体方法是得到一组 n 个看似产生于正态分布的数据，然后用最小的数

据替换原始数据中最小的一个，用第二小的数据替换原始数据中第二小的一个，依此类推. 因此用伪正态随机数中赋秩 k 的数代替原始观测值中赋秩 k 的观测. 注意在这个检验中我们只需要知道原始观测值的秩，就可以完成这种替代. 因此这是一个秩检验的统计过程. 伪随机样本可以从有关随机数表上获取，如一本名为 *A Million Random Digits with* 100,000 *Normal Deviates*(Rand Corporation,1955)的书上就有这样的表，或者设计具体的计算机程序生成. 这样的数据叫做“随机正态离差”，尽管它们并不是真正意义上的随机，它们只是经过谨慎考虑产生的似乎服从于标准正态分布的近似随机样本.

例如，我们先前使用的正态得分的数据为，7.3,7.7,9.2,12.0,26.4，我们由表中获取一组 5 个正态离差为：0.026，-1.388,2.388,1.066，-0.173，用它们中最小的值 -1.388 替代 7.3，次小的值 -0.173 替代 7.7，等等. 从这点上来看，这些新数据的使用与正态得分和秩的应用很相似. 当然，有的人在分析相同的数据时可能使用不同的 5 个数，这就会导致分析的结论略有不同（有时甚至有较大的区别）. 最糟糕的情况是两个人用同一种分析同一种数据得到矛盾的结论. 正是因为这个原因，在现实的分析过程中很少用这样的分析. 但是由于它的 A. R. E. 与正态得分的检验相同，而且它的精确分布与参数检验相同，所以人们从理论的观点对于研究它很感兴趣.

Bell 和 Doksum(1965)详细地阐述了关于随机正态离差的原则，而早先的分析出现在 Durbin(1961),Fraser(1957)以及 Ehrenberg(1951)的文章中.

期望正态得分

一种看待正态离差方法的观点是，认为用服从正态分布的次序统计量 $Z^{(i)}$ 来代替实际的次序统计量 $X^{(i)}$. 下面我们所考虑的得分，即用 $Z^{(i)}$ 的均值 $E(Z^{(i)})$ 代替次序统计量本身. 这些期望正态得分是已经定义好的数，在一些表中可以查到，如 Fisher 和 Yates(1957),Pearson 和 Hartley(1962)以及 Owen(1962). 因此在用 $E(Z^{(i)})$ 代替 $Z^{(i)}$ 404
(如随机正态离差）时，由 $Z^{(i)}$ 的变异性引起的麻烦即被消除. 这一类方法也只依赖于观测值的秩，因此也是一个秩检验. Fisher 和 Yates（1957）建议用这些精确的得分代替原始数据，然后对这些期望正态得分数据应用参数方法，从而得到一个非参数方法. 这些方法的 A. R. E. 与正态得分方法以及随机正态离差方法相同. Bradley(1968)对于这种变化给出了一个较为复杂的表达.

□**理论**　我们这里使用前几节中寻找检验统计量精确分布的方法，只不过有些细微的修改. 其一就是我们不使用秩 1,2,3,…，取而代之的是一些数，记它们为：$a(1)$，$a(2)$,$a(3)$,…,其中，$a(i)$表示正态得分，期望正态得分或者由独立于数据所获取的其他任一数集. 其二是所得的新数据有着与秩不同的均值和方差，我们需要确定这些均值和方差.

我们用下面的例子来说明寻找精确分布的方法. 如在 Mann-Whiteny 检验中，给

定两个独立样本 $X_1,X_2,\cdots,X_n$ 和 Y_1，$Y_2,\cdots$，Y_m，样本容量分别为 n 和 m. 在分布相同的零假设成立时，赋给变量 X_i 的秩是它在 1 到 $n+m$ 中等可能取到的任一个，因此 X_1 的任一得分在 $a(1),a(2),a(3),\cdots,a(m+n)$ 中也是等可能的. 依次可以类推至 $X_2,X_3,\cdots,X_n$ 以及 Y_1,Y_2,Y_3 等等. 因此，可以看出给 X 赋 n 个秩，共有 $\binom{m+n}{n}$ 种可能的赋秩方式，而每一种方式都是等概率的，概率为 $1/\binom{m+n}{n}$，这也意味着在 $a(1)$ 到 $a(m+n)$ 中，对 X 赋以 n 个得分，共有 $\binom{m+n}{n}$ 种可能的赋分方式，每一种方式出现的概率为：$1/\binom{m+n}{n}$. 这使得基于赋给 X(或 Y)的得分或秩的统计量的零分布可以用前面所述的计数方法来求得.

具体地，令 $n=2,m=3$，我们所用的得分为正态得分，

$$a(1) = -0.9661$$
$$a(2) = -0.4316$$
$$a(3) = 0.0000$$
$$a(4) = 0.4316$$
$$a(5) = 0.9661$$

405 则对 X_1,X_2，它们可能的秩、相应的得分以及得分和，如下表.

$(R(X_1),R(X_2))$	得分 (A_1,A_2)	和	概率
(1, 2)	(−0.9661, −0.4316)	−1.3977	0.1
(1, 3)	(−0.9661, 0.0000)	−0.9661	0.1
(1, 4)	(−0.9661, 0.4316)	−0.5345	0.1
(1, 5)	(−0.9661, 0.9661)	0.0000	0.1
(2, 3)	(−0.4316, 0.0000)	−0.4316	0.1
(2, 4)	(−0.4316, 0.4316)	0.0000	0.1
(2, 5)	(−0.4316, 0.9661)	0.5345	0.1
(3, 4)	(0.0000, 0.4316)	0.4316	0.1
(3, 5)	(0.0000, 0.9661)	0.9661	0.1
(4, 5)	(0.4316, 0.9661)	1.3977	0.1

因此我们可以求出得分和的分布函数. 类似地，我们可求出本节所涉及的任一统计量的分布函数. 但是我们不打算列出精确分布的表格，而是使用它们的渐近分布. □

在两样本的情况下，为了找到秩和的均值和方差，我们可以使用 5.3 节中提到的方法，该节中得到的结论可以直接用于本节的情况，用（5.3.18）式求均值，（5.3.24）式求方差. 使用得分代替秩的情况更加复杂，Hajek 和 Sidak(1967)对此问题进行了详细的讨论，同时描述了如何选取特殊情况下的最佳得分，并提供了一套

完整的理论，尽管这一理论已经超出了本书的范围，但是 Hajek 和 Sidak 的书是值得一读的，并推荐给大家.

关于随机正态数的讨论，我们可以参考 Marsaglia(1968)或者 Lewis(1975)，这也是众多有关这方面参考书中的两本. Jogdeo(1966)证明了对于某些确定的备择假设，随机正态离差的相对效率小于 1. Ramsey(1971)考查了基于两样本检验的小样本功效，而 Raghavachari (1965b), Thompson, Govindarajulu, Doksum (1967), Bhattacharyya(1967), Stone(1968)以及 Gokhale (1968)考虑了检验的大样本效率. Bradley, Patel 和 Wackerly(1971)在多元情况下讨论了这些检验的一些变化；Johnson 和 Mehrotra(1972)对删失数据讨论了这些检验的一些变化；Pirie 和 Hollander(1972)讨论了在随机区组设计中有序备择假设下的这些检验的变化问题. 有关这些方法的多角度分析可以参考 Lehmann(1975)或者 Hogg(1976).

习题

1. 在习题 5.2.1 中用正态得分代替秩，并比较两种方法所得的结果. 406
2. 在习题 5.2.3 中用正态得分代替秩，并比较两种方法所得的结果.
3. 在习题 5.7.1 中用正态得分代替秩，并比较两种方法所得的结果.
4. 在习题 5.7.3 中用正态得分代替秩，并比较两种方法所得的结果.
5. 在习题 5.3.1 中使用 Klotz 检验对数据进行分析，并比较两种方法所得的结果.
6. 在习题 5.3.2 中使用 Klotz 检验对数据进行分析.
7. 在习题 5.4.1 中运用正态得分计算相关系数（根据（10）式）. 如何比较这个系数和 Spearman 及 Kendall 系数的大小？比较 $\rho\sqrt{n-1}$ 和由表 A1 所得的分位数，对独立性假设进行显著性检验. 比较这一结果与习题 5.4.1 的结果.
8. 在习题 5.4.3 中运用正态得分进行趋势性检验. 比较 $\rho\sqrt{n-1}$ 和由表 A1 所得的分位数进行显著性检验，并比较这一结果与习题 5.4.3 的结果.

思考题

1. 对 $n=5$，求由（6）式定义的统计量的精确分布.
2. 对 $n=2, m=3$，求由（8）式定义的 Klotz 统计量的精确分布.
3. 用随机数表或产生正态随机数的计算机程序获得 34 个随机正态离差，将它们从小到大排序后代替例 5.10.1 中的正态得分. 如何比较 Bell-Doksum 方法的结果与 van der Waerden 及 Kruskal-Wallis 检验的结果？

5.11 Fisher 随机化方法

在前面的几节中，我们介绍了一些获得非参数检验的方法，每一种方法都用一个得分集合 $a(1)$ 到 $a(N)$ 代替秩 1 到 N. 常用的得分包括标准正态分布的分位数，或是一些产生于伪正态分布的随机样本，或者来自于标准正态分布次序统计量的期望.

我们已经提到过任何数都可以作为得分，但是有些类型的数作为得分，对于某些特定的备择假设会产生更高的功效.

假定我们要寻找到一组“较好”的得分代替秩1到N，并决定使用样本中实际的数据，这些数据很方便使用，因为它们正好有N个，而且容易获取. 如
407 同前几节中我们使用的正态得分一样，我们用这些数作为得分来进行非参数检验，但是选择这个得分是否比选择正态得分有更高的功效呢？根据Lehmann和Stein(1949)，Hoeffding(1952)以及其他人的研究显然是这样的，他们还发现在某些情况下，这些方法的A. R. E. 与最有功效的参数检验相比是1.0. 因此在这种情况下，这些得分不仅比正态得分，也比其他一些检验更受欢迎. 为什么呢？如果是这样的话，在假设检验中为什么我们不用数据直接代替得分呢？

但是，这个方法的最主要的缺点在于，它使得检验的操作变得冗长. 由于对每一个检验这些得分都是不同的，所以不可能构造临界域的表或者是检验统计量零分布的分位数. 因此每一次我们使用这种检验，就需要根据观测到的数据具体确定临界域. 每一个不同的样本就意味着不同的得分集合以及不同的临界域. 即使在条件容易满足，检验统计量的渐近分布为一个标准分布，如正态分布或χ^2分布的情况下，用这种渐近分布作为近似分布对于某些得分来说也不一定是准确的. 当得分是秩，正态得分或者期望正态得分时，我们至少知道得分是什么，也知道近似分布的精确性，并且在那些情况下，渐近分布可以作为很好的近似. 但是当得分集合随着样本而变化时，这就很难确定渐近分布近似的精确性了. 因此简而言之，我们也许经过努力得到精确的p-值（当样本量不小时需要相当大的努力）. 寻找近似p-值的方法是有的，但是并不一定精确. 由于用*StatXact*可以求得精确的p-值，这也就从本质上去除了这个不利影响.

第二个缺点是对本节中的随机化检验，它们缺少相对功效. R. L. Iman和本书作者未发表的模拟研究表明，对于很多分布，Fisher随机化检验的功效介于秩检验和参数检验之间. 总的来说，对于非正态的重尾分布或那些存在奇异值的数据，一个通常的秩检验，如Kruskal-Wallis检验倾向于比Fisher随机化检验更有功效.

一些学派考虑所研究的样本实际上不是来自于假设总体中的随机样本，而是总体自身. 此时测量集合就是我们感兴趣的总体，本节中讨论的随机化检验能够而且应当决定集合子群影响的存在与否. 更完整的哲理性表达参见Kempthorne和Doerfler(1969).

用数据本身作为得分的想法是由Fisher（1935）引出的，结果检验就是传统的随机化检验. 尽管我们的表达可能使人们认为随机化检验是第3代的非参数检验，排
408 在秩检验和其他得分检验之后，其实随机化检验在时间上是早于其他检验的. 随机化检验可以用于我们描述过的任何一个秩检验可以用到的地方. 我们将详细地给出两个独立样本及其配对样本的随机化检验方法，并给出相应的例子，以便说明如何使用这些检验. 首先一个检验就与5.1节中的Mann-Whitney检验相似.

▶两个独立样本的随机化检验

数据 由两个样本容量分别为 n 和 m 的随机样本 $X_1, X_2, \cdots, X_n$ 和 $Y_1, Y_2, \cdots, Y_m$ 组成.

假定条件

1. 两个样本是来自于它们各自总体分布的随机样本.
2. 除了每个样本中的各个变量相互独立外，两个样本之间也是相互独立的.
3. 度量尺度至少是区间的.

检验统计量 检验统计量 T_1 是 X 观测值的和

$$T_1 = \sum_{i=1}^{n} X_i \tag{1}$$

零分布 将 X 和 Y 混合成一个数据集合，从中抽取出 n 个数，在零假设成立的前提下，每一种 n 个数的组合方式都是等概率的，考虑所有组合的可能性即得到零分布. 因为不同的实际情况得到的 X 和 Y 是不同的，因此无法构造数学表，也很难确定相应的近似分布.

假设 我们只分析双边检验，单边检验与 5.1 节中提到的 Mann-Whitney 检验很相似.

$$H_0: E(X) = E(Y)$$
$$H_1: E(X) \neq E(Y)$$

如果统计量 $T_1 > w_{1-\alpha/2}$ 或者 $T_1 < w_{\alpha/2}$，则以水平 α 拒绝 H_0，其中分位数 w_p 用如下方法求得.

将观测到的 X_i 和 Y_j 值视为一个含有 $m+n$ 个数的数组，并从中取出 n 个样本，共有 $\binom{m+n}{n}$ 种可能的选择情况. 为了找到 p 分位数 w_p，考虑 $\binom{m+n}{n}$（p）个最小的次序和，即 T_1，其中最大的 T_1 就是 w_p.

如上述，如果 $\binom{m+n}{n}(p)$ 不是一个整数，就取大于它的第一个整数. 如果 $\binom{m+n}{n}(p)$ 是一个整数，w_p 就是最大的 T_1 的平均，而 T_1 是从所考虑的 $\binom{m+n}{n}(p)+1$ 种选择中得到的. 409

通过计算，从 $m+n$ 个数中选取 n 个数使它们的和小于（或是大于，如果观测的 T_1 处在右边）或等于所求得的 T_1 方式的个数，即可得到 p-值. 因为这是一个双边检验，所以我们用刚刚计算出的数值乘以 2，然后需要除以 $\binom{m+n}{n}$，即是 p-值.

计算机辅助 如果样本的容量不大，我们可以考虑所有可能出现的置换情况，然后通过 *StatXact* 求得该检验或其他检验的 p-值，对于置换总数太大的情况，可以通过随机选取其中足够多的置换来估计 p-值. ◀

例 5. 11. 1

假设随机样本 X_i 分别为 0,1,1,0, -2；随机样本 Y_j 分别为 6,7,7,4, -3,9 和 14.

零假设为：

$$H_0: E(X) = E(Y)$$

相应的备择假设：

$$H_1: E(X) \neq E(Y)$$

对两个独立样本进行随机化假设检验，$\alpha = 0.05$.

一个样本容量为 5，另一个为 7，所以从总共 12 个数中选取 5 个数的方式共有 $\binom{12}{5} = 792$ 种，因为（792）（0.025） = 19.8，所以我们需要找到 20 组 T_1 的次序最小值，从而求出 $w_{0.025}$. 这些数的组合以及相应的 T_1 值如下.

联合观测

	组	-3	-2	0	0	1	1	4	6	7	7	9	14	观测的 T_1 ($T_1 = \Sigma X$)
	1	X	X	X	X	X	Y	Y	Y	Y	Y	Y	Y	−4
	2	X	X	X	X	Y	X	Y	Y	Y	Y	Y	Y	−4
	3	X	X	X	Y	X	X	Y	Y	Y	Y	Y	Y	−3
	4	X	X	Y	X	X	X	Y	Y	Y	Y	Y	Y	−3
	5	X	X	X	X	Y	Y	X	Y	Y	Y	Y	Y	−1
	6	X	Y	X	X	X	X	Y	Y	Y	Y	Y	Y	−1
	7	Y	X	X	X	X	X	Y	Y	Y	Y	Y	Y	0
	8	X	X	X	Y	X	Y	X	Y	Y	Y	Y	Y	0
	9	X	X	X	Y	Y	X	X	Y	Y	Y	Y	Y	0
	10	X	X	Y	X	X	Y	X	Y	Y	Y	Y	Y	0
	11	X	X	Y	X	Y	X	X	Y	Y	Y	Y	Y	0
	12	X	X	Y	Y	X	X	X	Y	Y	Y	Y	Y	1
	13	X	X	X	X	Y	Y	Y	X	Y	Y	Y	Y	1
410	14	X	X	X	Y	X	Y	Y	X	Y	Y	Y	Y	2
	15	X	X	X	Y	Y	X	Y	X	Y	Y	Y	Y	2
	16	X	X	Y	X	X	Y	Y	X	Y	Y	Y	Y	2
	17	X	X	Y	X	Y	X	Y	X	Y	Y	Y	Y	2
	18	X	Y	X	X	X	Y	X	Y	Y	Y	Y	Y	2
	19	X	Y	X	X	Y	X	X	Y	Y	Y	Y	Y	2
	20	X	X	X	X	Y	Y	Y	Y	X	Y	Y	Y	2

由此得到最大的 T_1 是

$$w_{0.025} = 2$$

尽管这是一个双边检验，但是没有必要求 $w_{0.975}$，因为此时观测到的 T_1 已经处于左边范围. 因为由数据得到 T_1 的观测为：

$$T_1 = \sum_{i=1}^{5} X_i = 0 + 1 + 1 + 0 - 2 = 0$$

它小于 $w_{0.025}=2$，所以拒绝 H_0. 事实上，可以在水平为

$$p\text{-值}=\frac{2(11)}{792}=0.028$$

拒绝 H_0，其中有 11 种可能的排列方法使得得到的值小于或等于 0. ■

我们前面介绍的随机化检验是一个很典型的一般随机化检验方法. 如果不使用 T_1，我们也可以应用两样本的 t 统计量对 20 种组合分别计算分布的两个尾部概率，就像我们在该例题中所操作的一样. 但是这实际上是不需要的，因为正如5.1 节中提到的一样，t 统计量是 T_1 的一个单调函数，所以得到最大 T_1 的这 20 种组合与得到最大 t 的组合相同. 我们之所以提出这一点，是为了使随机化检验能推广到如单向表、双向表、相关性检验等情形，变得更加明显. 我们常常用一个统计量或者比较容易计算这个统计量的单调函数来决定数据的极端排列情况，及其所用统计量的临界域. 由于在计算过程中往往存在着一些困难，所以有时我们不求临界域，只求 p-值，特别是对于那些 p-值接近于零，且只需要考虑数据少数排列的情况.

对于配对的随机化检验与一般的随机化检验略有不同，下面我们将具体加以说明. 这种检验与 5.7 节中的 Wilcoxon 符号秩检验相仿. 411

▶配对的随机化检验

数据 数据由 n' 个二维随机变量 $(X_1,Y_1),(X_2,Y_2),\cdots,(X_{n'},Y_{n'})$ 组成，除去 (X_i, Y_i) 相等的组合即 $X_i-Y_i=0$ 的项，记剩余组合的个数为 n，并记非零差 X_i-Y_i 为 $D_1,D_2,\cdots,D_n$.

假定条件

1. 每一个 D_i 的分布是对称的.
2. D_i 是相互独立的.
3. D_i 有相同的均值.
4. D_i 的度量尺度至少是区间的.

检验统计量 检验统计量 T_2 为所有正差的和

$$T_2=\sum D_i \qquad \text{其中求和仅对 } D_i>0 \text{ 的项} \tag{2}$$

零分布 通过计算出 D_i 的正负号所有可能的排列，可以得到 T_2 的零分布. 在零假设下，每一种正负号的排列都是等可能的. 因为 T_2 的值依赖于 D 的值，即 T_2 随着观测值的变化而变化，因此我们不可能构造分位数表，或是确定可能近似分布的确切程度.

假设 我们只分析双边检验，单边检验可与 5.7 节中的 Wilcoxon 符号秩检验方法相比较得到.

$$H_0: E(D)=0 \qquad (\text{即 } E(X)=E(Y))$$
$$H_0: E(D)\neq 0 \qquad (\text{即 } E(X)\neq E(Y))$$

如果统计量 $T_2>w_{1-\alpha/2}$ 或者 $T_2<w_{\alpha/2}$，则以水平 α 拒绝 H_0，其中分位数 w_p 的用如下方法求得.

我们只考虑 D_i 的绝对值 $|D_i|$，而不考虑原来它们是正还是负. 那么，共有 2^n 种方法给这些所得绝对值加上正负号，例如，我们把正号加在所有的 $|D_i|$ 上，把正号只加在 $|D_1|$ 上，而把负号加在 $|D_2|$ 到 $|D_n|$ 上，等等. 为了找到 p 分位数 w_p，其中 $0 \leqslant p \leqslant 1$，首先找到 $(2^n)(p)$ 种符号的排列可以得到次序最小的 T_2（正的绝对差）.
412 [如果 $(2^n)(p)$ 不是一个整数，就取大于它的第一个整数.] 由此得到最大的 T_2 值就是零假设下的 p 分位数 w_p. [如果 $(2^n)(p)$ 是一个整数，则分位数是 T_2 的最大值的平均值，一般考虑是 $(2^n)(p)$ 与 $(2^n)(p)+1$ 种排列时两个最大 T_2 值的平均].

上述求 w_p 的方法在理论上可以应用于所有的 $0 \leqslant p \leqslant 1$，但是在实际情况下，我们往往只需要求较小的 p，如 $p=\alpha/2$. 对于 p 值较大的情况，我们可以用如下关系求得

$$w_{1-\alpha/2} = \sum_{i=1}^{n} |D_i| - w_{\alpha/2} \tag{3}$$

（3）式的结论是显而易见的，因为只须将每一个获得次序较小的 T_2 值的符号改变成相反号（用正号代替负号或用负号代替正号），我们就可以得到相应次序较大的 T_2. 用后一个 T_2 的值（所有正的 $|D_i|$ 的和）加上前一个 T_2 的值（所有负的 $|D_i|$ 的和）就得到了 $\sum_{i=1}^{n} |D_i|$，即（3）式中的关系.

通过计算得到小于（或者是大于，如果 $T_2 > \frac{1}{2}\sum_{i=1}^{n}|D_i|$）或等于 T_2 值的这种符号排列方法的个数，其中 T_2 由数据计算而得，我们就可以求得 p-值，因为只要将所得的个数乘以2，然后再除以（2^n），即为 p-值.

计算机辅助 *StatXact* 可通过考虑所有可能出现的置换方式求得该 Fisher 随机化检验的精确 p-值. 对于置换方式太多的情况可以通过随机选取其中足够多的样本来估计 p-值. ◀

例 5.11.2

假设从8组配对的数据中计算出的差为：$-16, -4, -7, -3, 0, +5, +1, -10$，去掉0后，我们得到，

$$D_1=-16, D_2=-4, D_3=-7, D_4=-3, D_5=+5, D_6=+1, D_7=-10$$

其中 $n=7$. 对于零假设

$$H_0: d_{0.50}=0$$

对备择假设

$$H_1: d_{0.50} \neq 0$$

在 $\alpha=0.05$ 水平下，使用随机化检验.

考虑最小次序"正的"绝对值和的4 [$(2^7)(0.025)=3.2$] 种符号排列方式，我们可以得到分位数 $w_{0.025}$. 如下给出，
413

符号的排列	$\sum$“正的”$\lvert D_i \rvert$
−16, −4, −7, −3, −5, −1, −10	$T_2 = 0$
−16, −4, −7, −3, −5, +1, −10	$T_2 = 1$
−16, −4, −7, +3, −5, −1, −10	$T_2 = 3$
−16, −4, −7, +3, −5, +1, −10	$T_2 = 4$
(−16, +4, −7, −3, −5, −1, −10 也给出了	$T_2 = 4$)

T_2 的最大值是 4，所以

$$w_{0.025} = 4$$

从（3）式我们得到

$$w_{0.975} = \sum_{i=1}^{7} |D_i| - w_{0.025} = 46 - 4 = 42$$

0 由数据得到检验统计量的值为：

$$T_2 = \sum \text{正的 } D_i = 5 + 1 = 6$$

因为得到检验统计量的值，既不小于 6，又不大于 42，因此接受 H_0.

p-值可从列出导致 $T_2 \leq 6$ 的符号排列中获得，除了上述已列出的 5 种，还有

符号的排列	$\sum$“正的”$\lvert D_i \rvert$
−16, +4, −7, −3, −5, +1, −10	$T_2 = 5$
−16, −4, −7, −3, +5, +1, −10	$T_2 = 5$
−16, −4, −7, −3, +5, +1, −10	$T_2 = 6$

共有 8 种符号排列的方式给出了统计量小于或等于 T_2 的值. 因为这是一个双边检验，8 种方式需乘以 2，从而，得到的 p-值是

$$p\text{- 值} = \frac{2(8)}{2^7} = \frac{16}{128} = 0.125$$

■

□**理论** 随机化检验背后的理论可以部分地通过求临界域的方法来解释. 例如，检验两个独立样本的时候，很明显，我们考虑每一种从 $n+m$ 个观测中选择 n 个 X 的方式具有等可能性. 剩下的需要解释为什么我们考虑的选择是等可能性的，以及为什么我们把观测值本身当作“样本空间”，下面加以解释. 414

我们所考虑的选择方式是等可能性的，这是因为零假设（包括假定条件）说明 X 和 Y 所有都是独立同分布的. 因此 X 不应该比 Y 有更低、或者更高、或者在中间的趋势. 任意给定一组 $n+m$ 个数，不管它们是否为观测，每一个含有其中 n 个数的子集都可能是 X 的 n 个值，就像任意其他含有 n 个数据的子集一样，因为这些数据不是来自 X 就是 Y，而这些数据子集上的概率也不取决于它们是来自 X 还是 Y. 现在，如果 X 与 Y 的分布不同，则数据是来自 X 还是 Y 就会有影响，但是对于为了得水平为 α 的临界域，我们限制随机变量是独立同分布的. 因此在直觉上我们认为所考虑的 X 的 n 个观测的选择方式是等可能的.

这也引出了第二个问题，“为什么我们将观测值本身作为样本空间”？我们前面已经解释过，任何 $m+n$ 个数据的集合可满足“等可能性”的要求. 但是在检验时，

我们需要确认所得到的 $m+n$ 个观测，就是这 $m+n$ 个数. 在秩检验中，使用的 $m+n$ 个数是从 1 到 $m+n$ 的整数，并且它们作为秩赋给观测值与观测之间建立了一一对应. 在这种情况下，我们使用观测本身作为数，这就消除了如何赋数到观测值上的问题，这一问题在秩检验中，当有结存在时，会使赋秩方法比较困惑. 通过把观测本身作为数来使用，可以很简单地将 n 个数字的一个选择确认为一个实际得到的数据. 然后借助于检验统计量，可以确认超过这个统计量所有极端值的选择方式、计数、并且用于计算 p-值.

临界域是由样本空间的个体子集所确定的，如那些与数据中观测值有相同数值结果的子集. 这些子集互不相容，覆盖了所有的样本空间（给定任何观测集，我们能求得这个样本空间子集的临界域），并且每一个子集具有一个与整个子集大小有关，水平为 α 的临界域，因此，组合所有临界域的总体水平也是 α，说明这是一个有效的检验.

两个独立样本检验和配对检验的主要区别是在配对检验中，对称性的假设用于说明改变代数符号而不改变概率，当它的分布关于零对称时，如果差 D_i 是 +6，那么出现 -6 的概率也是同样的，再说，使用什么数是没有关系的. Wilcoxon 检验使用秩，随机化检验使用观测本身作为数，以便我们可以对于一组实际获得的数据很容
415 易确认符号的排列. □

Fisher(1935)讨论过配对的随机化检验，两个独立样本的随机化检验是 Pitman (1937/1938)提出的，同时他还阐述了相关性的随机化检验和方差分析检验.

Chung 和 Fraser(1958)提出了多维数据的随机化检验. Welch(1937), Scheffé (1943), Moses(1952), Smith(1953)和 Kempthorne(1955)在他们的文章中给出了随机化检验的进一步讨论. Sen(1967b)在他的文章中讨论了多样本置换检验. Colllier 和 Baker(1963, 1966)和 Cleroux(1969)讨论了有用的检验统计量分布的逼近. 讨论 Fisher 随机化检验的其他文章包括 Tsutakawa 和 Yang(1974), Oden 和 Wedel(1975), Boyett 和 Shuster(1977)以及 Soms(1977).

习题

1. 一个轮胎公司对 10 名顾客进行跟踪研究，这 10 名顾客是从 3 年前在他们公司买新轮胎的顾客中随机选择的，问他们遇到过多少次（不管任何原因造成的）轮胎故障，如钉子，阀漏，等等. 这个研究限制在两个长寿命轮胎线上，称为 A 品牌和 B 品牌. 下面是数据结果.

顾客	A品牌	B品牌
1	0	3
2	2	5
3	0	1
4	1	4
5	2	3

用 Fisher 随机化检验方法来得到检验零假设：轮胎故障等可能，单边备择假设为：A 品牌的轮胎故障更少的精确 p-值.

2. 取 8 个成人的随机样本，问他们第一次约会的年龄. 3 位男士回答是 15,17,16 岁，而 5 名女士则回答 12,14,15,10 和 12. 检验假设：两种性别的平均年龄是相同的，备择假设是：女孩子第一次约会的年龄倾向于比男孩子更年轻.
3. 每小时观测两名售货员服务的顾客数，记录差 $Y_i - X_i$，其中 Y_i 和 X_i 分别代表每名售货员服务的顾客数. 检验：差 $Y_i - X_i$ 的中位数是否可以认为是零，观测的差是 +7, +3, +2, +8, −2, +3, +4 和 −1.
4. 2 名高速公路巡警检查他们 7 天所开的交通罚单数，Y_i 和 X_i. 配对观测 (X_i, Y_i) 是 (17, 14), (15,14), (12,15), (9,7), (17,16), (18,18) 和 (14,10)，问 $Y_i - X_i$ 的中位数是零吗？ 416

思考题

1. 某人建议在两个独立样本的随机化检验中，从所有观测里减掉一个常数，使得计算更简单，如在习题 2 中，在分析数据前从每个观测中减掉 10. 这会影响检验结果吗？请解释. 从观测中除掉一个非零常数会影响结果吗？
2. 配对随机化检验的结果会受从所有的观测中减掉一个常数或除掉一个非零常数的影响吗？请解释.
3. 在相关性的随机化检验中，如在秩相关性检验中，临界域的确定假设了每对 X 和 Y 是等可能的，其中数据是由二维样本 $(X_i, Y_i), i = 1, 2, \cdots, n$ 组成. 解释如何求得在零假设 X 和 Y 独立下，检验统计量 $T_3 = \sum_{i=1}^{n} X_i Y_i$ 的 p 分位数 w_p.

5.12 秩变换的讨论

这一章所讲述的大部分非参数方法是一些对数据应用秩变换方法的例子（即用秩替换数据），然后再使用通常的参数方法，此时针对秩而不是对于真实数据. 我们在 5.4 节提到了秩变换最明显的应用，就是把通常的乘积矩相关系数，即熟知的 Pearson γ 应用到秩上，可得到 Spearman ρ. 在 5.6 节中，我们对秩而不是对原始数据求出了通常的最小二乘回归线.

其他的非参数方法，如 Mann-Whitney 检验，Wilcoxon 符号秩检验，Kruskal-Wallis 检验，就不是很明显的秩变换. 让我们下面考查一下 Mann-Whitney 检验.

5.1 节中的 Mann-Whitney 检验考虑了两个随机样本，看它们所来自的两个总体是否具有相同的均值. 如果总体是正态时，则两样本的 t 检验是最有功效的检验，它是通过比较由（5.1.17）式给出的统计量 t 与服从自由度为 $N-2$ 的学生 t 分布的分位数来进行的，其分位数由表 A21 给出.

非参数的 Mann-Whitney 检验没有总体是正态性的假设，它是通过比较由（5.1.2）式给出的统计量 T_1，与作为其精确分布逼近的正态分布来进行的.

如果我们比较由秩构造的 t 统计量（（5.1.17）式）而不是由数据构造的统计量，结果将如何呢？结果显示这一方法等价于 Mann-Whitney 检验，只是用了不同的

逼近分布. 换句话说，令 t_R 为由（5.1.17）式计算出的基于秩而不是观测本身的 t
417 统计量，则这个基于 X 和 Y 秩计算的两样本统计量是

$$t_R = \frac{T_1}{\sqrt{\frac{N-1}{N-2} - \frac{1}{N-2}T_1^2}} \tag{1}$$

注意，当 T_1 增大时，t_R 也增大，T_1 减小时，t_R 也减小. 这就意味着如果检验因为 T_1 太大或者太小而拒绝零假设，那么检验也同样会因为 t_R 太大或者太小而拒绝零假设. 所以两种检验，即 Mann-Whitney 检验和秩变换方法确实是等价的，只要将 T_1 的 0.95 分位数代入（1）式就得到 t_R 的 0.95 分位数. 但结果与从自由度为 $N-2$ 的 t 分布中求得的 0.95 分位数（见表 A21）不是完全一样，而后者是当精确值不知道时，对精确值的一个很好的近似. 所以基于 T_1 和 t_R 的检验是等价的检验.

在 5.2 节中，（5.2.3）式给出了 Kruskal-Wallis 检验统计量 T，（5.2.19）式给出了在单因素方差分析中使用的统计量 F，基于观测的秩计算的 F 统计量是一个 T 的函数：

$$F_R = \frac{T/(k-1)}{(N-1-T)/(N-k)} \tag{2}$$

并且，因为 F_R 随着统计量 T 的上升而上升，T 的下降而下降，所以秩变换方法等价于 Kruskal-Wallis 检验方法（见思考题 5.2.5）.

我们现在来看 Wilcoxon 符号秩检验，这个检验应用到一个差的随机样本 D_1，$D_2,\cdots,D_n$，用于检验均值相等的零假设，即 $E(D_i)=0$. 通常的参数方法需要假设 D_i 是来自于正态分布的随机样本，当由（5.7.18）式给出的 t 统计量过大或者过小时拒绝零假设，称它为单边的 t 检验，我们早在 5.7 节中提到过它. 对于 Wilcoxon 符号秩检验，用符号秩（细节见 5.7 节）R_1 到 R_n 代替 D_i，并且当由（5.7.5）式给出的 T 统计量过大或者过小时拒绝零假设. 秩变换方法建议，通过使用符号秩计算统计量 t 来构造一个新的检验，但是这个检验实际上并不新，因为基于符号秩的一样本 t 统计量，即 t_R 仅是一个关于 Wilcoxon 符号秩统计量 T 的一个函数，表示如下（见思考题 5.7.3）：

$$t_R = \frac{T}{\left(\frac{n}{n-1} - \frac{1}{n-1}T^2\right)^{1/2}} \tag{3}$$

而且，大的 T 值对应于大的 t_R 值，小的 T 值对应于小的 t_R 值，所以两个检验是等
418 价的.

这些例子以及本章中的其他例子，如 5.5 节中关于斜率的检验，Friedman 检验，以及 Durbin 检验，都表达了一种思想，即应用通常的参数检验统计量以及对应观测的秩来获得一个非参数方法，在大多数情况下，它具有较高的效率. 当然，对每一情形，以这种方法给观测赋秩的妙处是在零假设下，每一种可能的赋秩方式都是等可能的. Worsley(1977)把这一技巧成功地利用到聚类分析，Shirley(1977)也利用这

一技巧对比了一种处理的增长剂量水平.

这些非参数检验等价于在秩上计算的参数检验，它们可以较容易地用计算机中计算参数检验设计的程序求出. 在没有非参数检验的计算机程序时，我们可以给数据简单地赋秩，然后对秩使用参数检验的方法. 参数检验能自动地对结进行修正，这就使得一般近似的 p-值比通常由正态分布或者 χ^2 分布得到的近似要好.

在这种赋秩方法不可能或者很困难的情况下，秩变换的原则依然有用. 既使不能得到非参数检验，但在试验设计和多元回归两个统计分析方面，秩变换方法是很有用的.

为了分析秩变换在试验设计中的应用，首先从最小到最大对所有的观测赋秩，然后对秩使用通常的方差检验，其结果是一个条件分布自由的方法.

也就是说，在秩检验中可以得到一个检验统计量的精确分布，但是对于不同的秩结构，分布是不同的. 即想找精确分布并不实际，所以在很多情况下，将使用大样本逼近，这与参数检验中所使用的 F 分布基本相同.

秩变换可以很好地处理没有交互作用的双向表（见 Iman, Hora，和 Conover，1984，以及 Hora 和 Iman,1988)，此时它比 Friedman 检验，Quade 检验以及参数的 F 检验要好. 但是，试图对交互作用使用秩变换方法进行检验却没有得到一致的结论，在一些情况下，它具有很好的稳健性和效率，参见 Iman(1974b), Conover 和 Iman (1976)以及 Pavur 和 Nath（1986)，但是在另一些情况下，Blair, Sawilowsky 和 Higgins(1987)却说明这一检验是不稳健的，效率也不高. Thompson(1991)作了一个理论上的研究，指出了秩变换检验在检验交互作用时的缺点，说明它不是一个有效的方法，不应当使用它. 但是 Mansouri 和 Chang(1995)用正态得分代替秩，并发现用正态得分变换方法检验交互作用没有问题，所以正态得分的转化可以修正 Thompson 发现的缺点.

在试验设计中，对于那些非参数检验不存在的情况下，一个值得推荐的方法就是用通常的方差分析分析数据，然后使用相同的方法分析秩变换数据. 如果两个方法给出了几乎相同的结果，那么通常方差分析的基本假设可能比较合理，通常的参 419
数分析方法也是有效的. 当两个方法给出了完全不同的结论，那么试验者可能需要认真观察一下数据，特别是那些离群值（与一般数据相比大得出奇的值）或者是特别不对称的分布. 这些数据中的失常现象会很大程度上改变显著水平，降低效率，但是基于秩的分析就不会受到同样大的影响. Crouse(1967)，Lemmer 和 Stoker (1967)，Crouse(1968)，Macdonald(1971)，Scheirer, Ray 和 Hare(1976)，以及 Hamilton(1976) 在试验设计中使用了秩变换方法. 更多近期的研究参见 Akritas(1990 和 1991).

在多元回归中，对每一个变量分别进行赋秩，就像在 5.5 节中提到的二元回归方法，然后在秩上应用通常的回归方法，结果得到一种稳健的回归分析方法，它不像一般的回归方法对于离群值或是非正态分布那样十分敏感. 因此，如前面提到的，

建议通过分析数据、分析秩，并从两个角度分别对结果进行解释．对于相关变量的预测在5.6节中已经进行了分析，通过由回归方程预测秩，以及在相关变量的已知值中使用差值的方法．Iman和Conover(1979)给出了使用这个方法的例子．

秩变换在判别分析中的应用使得得出的方法很简单，并且在给观测分类时很有效．简单地说，分别赋秩给每一个变量，基于秩计算线性判别方程和二次判别函数．Conover和Iman(1978b,1980)给出了这个方法一个更详细的讨论，并且用蒙特卡罗方法进行了大量功效比较．

统计学的其他领域也可以更多地运用秩变换方法，这些方法通常不是分布自由的．但是当标准方法的假设不合理时，它们比标准过程更加稳健而且经常更有效．参考Hettmansperger和Mckean(1978)对使用秩进行的一般性讨论．其他稳健方法，不一定是基于秩的，现在也受到了很大的关注．有关这些稳健方法的一些重要参考文献包括Huber(1972)和Hogg(1977)．Labovitz(1970)和Allan(1976)也讨论了这些方法，Kim(1975)的文章列出了许多这方面的参考文献．

5.13　第5章复习题

1. 州高速公路管委会希望买一种比较好的油漆，喷刷高速公路上的标志线，最后确定要在两种品牌的油漆中选一种．在高速公路的一段涂上20条条纹，其中用品牌A涂10条，用品牌B涂10条，顺序是随机的．并且6个月后分别检查这些条纹，并根据磨损程度排
420 序．得到的结果如下：

	秩
品牌A	2, 3, 4, 6, 8, 9, 10, 12, 13, 14
品牌B	1, 5, 7, 11, 15, 16, 17, 18, 19, 20

品牌A和B的差别是否显著？

2. 在珠宝商店中金戒指有两种不同尺度测重，尺度A是电子测量，尺度B是机械式天平测量．为了看用尺度B测量的重量是否高于尺度A的测量，我们用两种尺度分别测量了7枚金戒指，结果如下．那么，用尺度B测得的重量比用尺度A测得的重量显著重吗？

戒指	尺度A	尺度B
1	22.6	22.9
2	13.8	14.3
3	19.0	19.1
4	26.5	26.4
5	24.9	25.2
6	16.0	16.4
7	23.3	23.4

(a) 使用Fisher随机化检验．

(b) 使用Wilcoxon符号秩检验．

3. 10位高尔夫球手同意在一次锦标赛中试验一种新球．从这10位选手中随机地选取5位使用新球，其他的5位使用旧球，4轮后结果如下，

新球得分	295	301	288	290	289
旧球得分	302	306	292	306	314

(a) 这些结果是否提供了使用新球会倾向于得分降低的证据?

(b) 在步骤 (a) 中还可以使用什么统计方法进行分析? 它们分别有什么优点和缺点, 包括你所使用的方法?

4. 一个球童在等人的过程中, 看见8名高尔夫球手在完成比赛后, 付钱给了他们各自的球童, 然后离开了. 他估计了每一个选手的年龄, 并记下了他们付给球童的报酬.

高尔夫球选手	1	2	3	4	5	6	7	8
年龄(估计的)	32	30	33	41	43	47	28	30
支付总额	10.00	11.50	9.00	12.00	16.00	17.00	8.75	10.50

(a) 这些数似乎说明年龄越高的人给球童的报酬越高吗?

(b) 在步骤 (a) 中还可以使用什么统计方法进行分析? 它们分别有什么优点和缺点, 包括你所使用的方法?

5. 两位赛马训练师要比较他们最新训练的5匹马的比赛结果, 看谁是训练赛马跑得更快的人. 下表给出了第一个训练师所训练的马跑1/4英里所需的时间. 421

马	1	2	3	4	5
训练前	26.3	24.1	27.6	25.3	26.8
训练后	23.3	22.0	24.1	22.8	23.0

下表给出了第二个训练师所训练的马跑1/4英里所需的时间.

马	1	2	3	4	5
训练前	25.4	26.2	24.0	26.0	27.7
训练后	23.6	23.9	21.8	23.6	25.7

检验假设: 两位训马师训练的马跑得一样快.

6. 为了检验进一步训练是否必要, 我们按顺序记录了同一匹马在10天内每天早晨跑1/4英里所需的时间, 结果如下

天	1	2	3	4	5	6	7	8	9	10
速度(秒)	22.2	22.8	21.0	21.4	22.4	21.9	22.0	22.6	21.8	21.1

这些数据是否说明这匹马的速度还可以提高?

7. 随机选择18个高中学生, 对他们进行品行评分 $X, X=10$ 代表满分; 成绩评分为 $Y, Y=20$ 代表20门课程中每一个都得到满意成绩.

学生	1	2	3	4	5	6	7	8	9
X	1.8	8.9	8.3	4.0	8.8	9.2	9.5	8.1	5.3
Y	11	17	16	10	16	17	20	16	11

学生	10	11	12	13	14	15	16	17	18
X	7.3	7.7	6.8	7.9	8.8	9.9	9.0	9.3	9.2
Y	14	15	12	14	17	20	18	19	18

(a) X 和 Y 之间有显著的正相关吗?
(b) 求出最小二乘回归直线.
(c) 求出最小二乘回归直线斜率的95%置信区间.
(d) 用秩回归估计回归曲线 $E(Y|X)$.
(e) 画一幅图，在图上给出数据点，最小二乘回归直线，及其用秩回归的单调回归曲线估计. 哪个回归估计看起来与数据更一致?

8. 随机选择男士和女士组成一个样本，其身高（英寸）如下所示，

男士	女士
$X_1 = 70.1$	$Y_1 = 62.2$
$X_2 = 67.8$	$Y_2 = 64.7$
$X_3 = 71.6$	$Y_3 = 65.3$

422

检验零假设：男士和女士的身高同分布，对备择假设：男士倾向于比女士更高些. 令检验统计量 T 为赋予 Y 的最大的秩，其中以高度增加为序，将秩1到秩6赋给男女联合样本.
(a) 求 T 在零假设下的概率分布.
(b) 求 T 在零假设下的概率分布函数，并绘图.
(c) 求一个合理的临界域，并找得显著水平.
(d) 用前面的检验来检验零假设.
(e) 用你所学过的或发明的其他非参数方法检验零假设.

9. 在新学年开始的时候，将一年级的学生随机地分成两个组. 第一组用一视同仁的方法教授阅读，即所有的学生在老师指导下，在同一时间内由一个水平提高到另一个水平. 第二组用因材施教的方法教授，即每一个人根据教材难度，并且在老师的辅导下，按照他或她自己的速度进行学习. 年底时，每一个学生都接受了一项阅读测验，结果如下.

第一组				第二组			
227	55	184	174	209	271	63	19
176	234	147	194	14	151	184	127
252	194	88	248	165	235	53	151
149	247	161	206	171	147	228	101
16	99	171	89	292	99	271	179

(a) 检验零假设：两种教学方法的效果没有区别，对备择假设：两个总体均值不相等.
(b) 检验零假设：两个总体方差相等，对备择假设：第二种方法的总体方差大于用统一方法教授阅读的总体方差.

10. 一所学校有121名学生，一个学期学生缺课人数汇总如下：

缺课次数	0	1	2	3	4	5	6	7	8	超过 8
学生数	54	32	10	4	5	5	3	0	1	7
(男生)	28	15	4	2	2	2	1	0	0	3
(女生)	26	17	6	2	3	3	2	0	1	4

(a) 讨论这个问题中“目标总体”和“样本总体”的概念.

(b) 女孩比男孩出勤记录更少吗?

(c) 女孩一般倾向于比男孩缺课更多吗?

11. 在一次地区级才艺竞赛中，要求7名考官给5个进入决赛的人排序. 赋秩从最好的1到最差的5，结果如下， 423

	表演者				
裁判	A	B	C	D	E
1	5	2	1	3	4
2	5	1	2	3	4
3	3	1	2	4	5
4	2	3	4	1	5
5	3	1	2	5	4
6	4	1	2	3	5
7	4	2	3	1	5

问秩是随机排列的零假设会被拒绝吗?

12. 某公司希望从5种不同类型的梳子中选出一种做营销推广的重点. 作为分析的一部分，他们选择了10个来自高中的女孩作为顾客，检验顾客对这些梳子的喜好. 给这10个女孩中的每一个这两种类型的梳子，让她们用一个月，然后要求她们汇报她们的喜好. 为了简单起见，将不同类型的梳子叫做A,B,C,D和E，结果如下.

Alice	A B 中更喜欢 B	Fawn	B D 中更喜欢 B
Betty	A C 中更喜欢 A	Greta	B E 中更喜欢 E
Charlene	A D 中更喜欢 D	Heather	C D 中更喜欢 D
Donna	A E 中更喜欢 E	Inga	C E 中更喜欢 E
Ellen	B C 中更喜欢 B	Jean	D E 中更喜欢 E

喜好有显著的差异吗? 如果有，哪些梳子有显著差异?

13. 几种普通股票在一个时期内的投资回报率可以这样计算，它由这个时期末每支股票的市场价格加上这个时期内所付的一些股息，然后用次结果除以这个时期开始时股票的价格而得到. 几种股票的投资回报率记录如下，共9个时期，每期3个月. 不同的股票看起来回报率有显著差别吗?

	股票				
时期	A	B	C	D	E
1	1.022	1.018	1.031	1.009	1.018
2	0.996	0.998	1.021	0.981	0.992
3	1.001	0.993	0.998	1.010	1.008
4	1.064	1.073	1.020	1.051	1.061
5	1.013	1.009	1.026	1.042	1.000
6	1.113	1.126	1.088	1.141	1.103
7	0.998	0.992	1.012	1.002	0.977
8	0.993	1.004	1.010	0.998	0.987
9	1.061	1.020	0.999	1.031	1.040

14. 在问题13相同研究的另一部分中，计算40支股票在9个时期、每期3个月中的总回报率. 所选的这40支股票代表4种不同类型的行业，每个行业有10支股票. 424

27个月的回报率

	行业类型			
股票	A	B	C	D
1	1.062	1.060	1.101	1.003
2	1.021	1.001	.981	1.067
3	1.000	1.124	1.173	1.084
4	1.316	.961	1.126	1.049
5	1.177	1.054	1.002	1.056
6	1.289	1.048	.964	1.012
7	1.405	1.113	1.142	1.008
8	1.566	1.147	1.226	1.051
9	1.304	1.067	1.184	1.058
10	1.111	1.073	1.098	1.042

这4种类型的股票看起来回报率有显著不同吗?

15. 一个乡村估价官记录了所有去年在某个乡镇附近卖掉的超过20英亩土地的价格，她将每笔销售的数据归结为两个变量：X = 到城市边界的距离，Y = 每英亩的价格.

	土地块														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X(英里)	12.1	4.8	13.9	1.6	17.4	7.5	19.9	21.8	2.4	5.8	2.3	12.8	25.6	8.8	7.3
Y(美元/英亩)	280	590	163	530	157	394	177	110	620	492	761	210	115	245	334

要求她为每块位于离城市边界4.4英里的土地提供一个公平的市场价格，只考虑前面的信息，问每英亩价格应该是多少?

16. 研究6个人，看他们早晨的休息心跳率是否高于晚上的，结果如下：

人	早晨	晚上
1	78	73
2	86	81
3	64	64
4	74	73
5	74	69
6	72	71

(a) 用正态得分型检验分析这些数据.

425 (b) 求斜率的90%置信区间，这里 Y = 早晨心跳，X = 晚上心跳.

17. 一个本地煤气站的随机样本有如下的雇员个数.

X 4, 5, 7, 12

一个连锁煤气站的随机样本有如下的雇员个数.

Y 9, 11, 15

(a) 用 X 的秩和，在 $\alpha = 0.05$ 下，检验 H_0：$\mu_x = \mu_y$，对 H_1：$\mu_x < \mu_y$. 求精确的 p-值(不用查表).

(b) 用 ven der Waerden 检验.

(c) 用 Fisher 随机化检验.

18. 4 个不同的承包人生产一种化学检测试剂. 假定由所有的承包人生产的所有试剂对有毒的气体的检测是相同的, 进行一项检验看是否有这种情况. 从每个承包人生产的许多试剂中随机抽取 10 种, 把这 40 种试验放入一个气体实验室中, 在给定试验条件下经过一段时间后再做比较. 试剂呈现出不同的颜色, 把这些颜色从粉色到深紫色赋秩如下:

承包人			
A	B	C	D
秩			
19	18	25	7
10	5	3	20.5
4	28	32	23
38	1	29	13
33	15	6	16
36	12	2	9
39	27	30	8
40	31	35	14
37	20.5	34	17
26	22	24	11

由不同承包人生产的试剂有差异吗?

19. 用计算机模型模拟进行红军和蓝军之间的 12 场战斗, 每次战斗中每支队伍的伤亡人数记录如下:

战斗	红军	蓝军
1	41	38
2	8	14
3	65	41
4	28	31
5	11	8
6	15	18
7	73	48
8	54	32
9	7	7
10	50	37
11	59	42
12	24	28

426

(a) 画出数据的散点图.

(b) 用 Spearman ρ 作为单调相关强度的度量.

(c) 用 Kendall τ 度量数据对之间协调性强度.

(d) 用单调回归方法估计, 当蓝军有 40 名伤亡人数时, 红军的平均伤亡人数.

20. 招收 8 名自愿者以检验在步枪上安装望远镜瞄准器的效果. 相信在步枪上安装望远镜瞄准器会提高射击目标的测试分数, 为了证明这个结论, 要求这 8 名自愿者每人用一把步枪以两种方式射击目标, 一种是安装望远镜瞄准器, 另一种是不安装, 而选这两种方式

的次序是随机的. 下面是数据结果：

	自愿者							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
安装望远镜瞄准器	96	93	89	88	85	83	80	77
没有安装望远镜瞄准器	92	92	89	96	82	79	80	78

安装望远镜瞄准器会得到较高的测试分数吗？

(a) 用一个仅基于秩的检验，求以下3种方式的 p-值：

　(1) 求精确 p-值.

　(2) 用没有连续性修正的正态逼近.

　(3) 用有连续性修正的正态逼近.

(b) 用基于正态得分的检验.

(c) 用 Fisher 型的随机化检验.

427 (d) 求在步枪上安装望远镜瞄准器所获得测试分数改进量的90%置信区间.

第 6 章 Kolmogorov-Smirnov 型统计量

导 言

在第 2 章中，我们介绍了经验分布函数，它是基于随机样本的函数，并且是可用来估计总体的分布函数. 如果我们想看两组或多组样本是否来自同一未知分布，很自然，可以比较它们的经验分布函数，看它们是否有相似之处. 确切地说，需要一个衡量两个或多个分布间差异的度量. Kolmogorov 和 Smirnov 提出和发展了用函数间的最大垂直距离作为分布相似性的一种度量的统计方法. 本章将讲述这个方法及其他运用这种思想的方法.

6.1 Kolmogorov 拟合优度检验

本章中我们先讲述 Kolmogorov（1933）提出的一种拟合优度检验. 这个检验可能是本章中最有用的检验，一部分原因是，它提供给我们一种替代4.5 节中对名义变量型数据所使用的拟和优度χ^2 检验的方法，使它可以适合于顺序类型的数据. 另一部分原因是，Kolmogorov 检验统计量使我们能够构造一个对于未知分布函数的“置信界”，这一点我们将在本节中解释. 428

检验拟合优度通常是考察一个来自某个未知分布的随机样本，检验其未知分布函数是否符合零假设为某个已知而具体的分布. 即，零假设具体指明了某个分布$F^*(x)$，可能是如图 6-1 所示的分布函数，也可能是一个可以画出其图像的数学函数. 通过某种方式将一组来自于某个总体的随机样本 $X_1, X_2, \cdots, X_n$ 与$F^*(x)$比较，来判断$F^*(x)$为这组样本的真实分布是否合理.

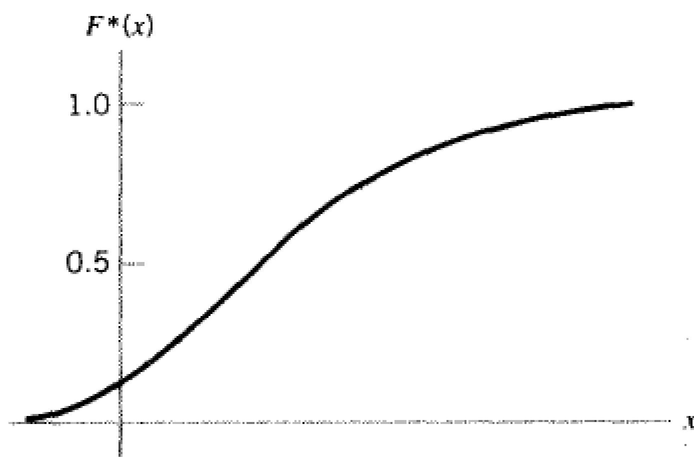


图 6-1 假设的分布函数

一种符合逻辑的办法就是，把随机样本的经验分布函数 $S(x)$ 与$F^*(x)$作比较.

由定义 2.2.1 所述，对任一 $x(-\infty<x<\infty)$，经验分布函数是 $X_i(i=1,2,\cdots,n)$ 中小于或等于 x 的比例. 在 2.2 节中我们知道，经验分布函数 $S(x)$ 是 X_i 的未知分布的一个很有用的估计，所以我们可以把 $S(x)$ 与所假设的分布函数 $F^*(x)$ 作比较，看它们是否吻合. 如果它们不能很好的吻合，我们可以拒绝零假设，并得出结论：此未知真实的分布函数 $F(x)$，不是由零假设中的 $F^*(x)$ 给定的.

但是，我们能使用什么类型的检验统计量作为 $S(x)$ 与 $F^*(x)$ 之间差异的度量呢？可想而知，一个最简单的度量就是用 $S(x)$ 与 $F^*(x)$ 在垂直方向上的最大距离，这是由 Kolmogorov（1933）提出的统计量. 亦即，如果 $F^*(x)$ 如图 6-1 所示，一组容量为 5 的样本抽自这个总体，把它的经验分布函数与 $F^*(x)$ 画在一起，如图 6-2 所示. 若 $F^*(x)$ 与 $S(x)$ 给定，那么这两者之间的最大垂直距离出现在 $S(x)$ 的第三阶以前. 这个距离在图 6-2 中大约为 0.5，因此，在这种情况下，Kolmogorov 统计量 T 就等于 0.5. 若 T 值超过表 A13 中所给出的值，则表明拒绝 $F^*(x)$ 作为未知真实分布 $F(x)$ 的合理逼近.

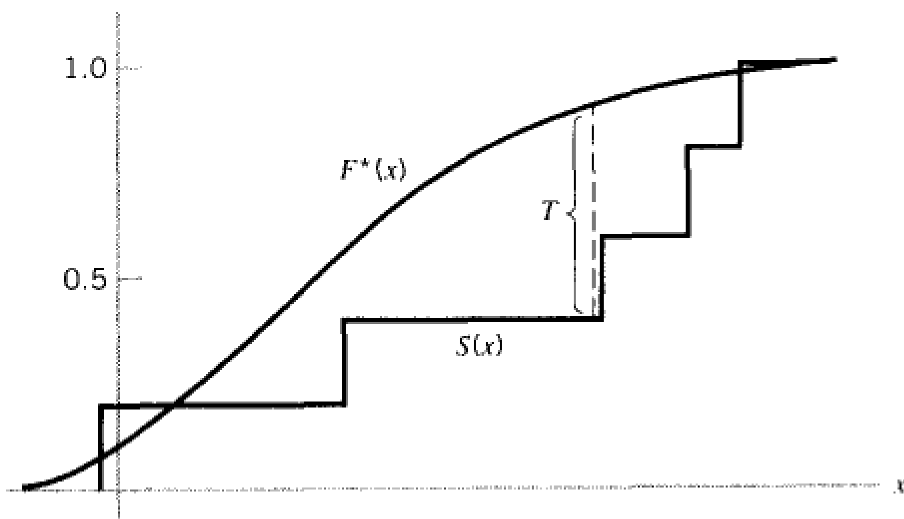


图 6-2 假设的分布函数 $F^*(x)$，经验分布函数以及 Kolmogorov 统计量 T

在小样本情形下，我们更愿意用 Kolmogorov 检验替代 χ^2 检验来拟合优度，即使
429 在小样本情形下，Kolmogorov 检验也是精确的，而 χ^2 检验则假设样本容量足够大，使得 χ^2 分布是检验统计量分布的一个比较好的近似. 哪一种检验更有效通常存在争论，但是一般感觉在多数情形下，包括顺序型数据，Kolmogorov 检验要优于 χ^2 检验. 想了解更多的这方面的比较，参看 Slakter（1965）.

本章的标题是“Kolmogorov-Smirnov 型统计量”，我们把 $S(x)$ 与 $F^*(x)$ 的最大垂直距离函数的统计量称为 Kolmogorov 型统计量，把两个经验分布函数最大垂直距离函数的统计量称为 Smirnov 型统计量. 本章仅考虑由分布函数间的垂直方向距离决定的统计量，既可以是经验分布函数，也可以是假设的分布函数.

▶Kolmogorov 拟合优度检验

数据 数据包含容量为 n 的随机样本 $X_1,X_2,\cdots,X_n$，它来自于某个未知分布 $F(x)$.

假定条件

1. 样本是随机样本.

检验统计量 令 $S(x)$ 是基于随机样本 $X_1, X_2, \cdots, X_n$ 的经验分布函数，$F^*(x)$ 是完全已知的假设分布函数．对于下列 3 个不同的检验 A，B，C，检验统计量的定义也有所不同． 430

A.（双边检验） 设检验统计量 T 为 $S(x)$ 与 $F^*(x)$ 的最大（记为"sup"）垂直距离，用式子表达如下：

$$T = \sup_x |F^*(x) - S(x)| \tag{1}$$

它读为"T 等于 $F^*(x) - S(x)$ 的绝对值对所有实数 x 的上确界"．

B.（单边检验） 记这个检验统计量为 T^+，它等于 $F^*(x)$ 位于 $S(x)$ 上方它们的最大垂直距离，也就是：

$$T^+ = \sup_x [F^*(x) - S(x)] \tag{2}$$

这与 T 类似，但要注意的是，这里我们只考虑 $F^*(x)$ 位于 $S(x)$ 上方那部分的最大垂直距离．

C.（单边检验） 记这个检验统计量为 T^-，它定义为 $S(x)$ 位于 $F^*(x)$ 上方它们的最大垂直距离，也就是：

$$T^- = \sup_x [S(x) - F^*(x)] \tag{3}$$

零分布 设 $F(x)$ 是连续的且零假设是正确的，那么 T^+ 与 T^- 的精确分布为：

$$G(x) = 1 - x \sum_{j=0}^{[n(1-x)]} \binom{n}{j} \left(1 - x - \frac{j}{n}\right)^{n-j} \left(x + \frac{j}{n}\right)^{j-1} \tag{4}$$

其中，$[n(1-x)]$ 是小于或等于 $n(1-x)$ 的最大整数，且 T^+ 与 T^- 的分布相同．$\sqrt{n}T^+$ 与 $\sqrt{n}T^-$ 的渐近分布为（当 $n \to \infty$ 时）：

$$H(x) = \lim_{n\to\infty} G\left(\frac{x}{\sqrt{n}}\right) = 1 - \mathrm{e}^{-2x^2} \tag{5}$$

T 的近似分布函数为：

$$P(T \leqslant x) \doteq [G(x)]^2 \tag{6}$$

这是由于，只有当 T^+ 与 T^- 同时小于 x 时，T 才小于 x．

$n \leqslant 40$ 时，双边检验中 T 精确的分位数以及单边检验中 T^+，T^- 近似的分位数在表 A13 中给出．近似的估计用于 $n > 40$．注意，所有的这些检验都只是右边的．"单边"或"双边"的假设根据个人的研究兴趣而定，检验统计量可以重新定义，使得 431
所有这 3 种检验都是右边的．

当 $F(x)$ 连续时，表 A13 是精确的，否则，所有这些分位数导致一个保守的检验（Noether，1967a）．当 $F(x)$ 为离散时，例 6.1.1 将描述一种用来找精确零分布的方法．

假设

A.（双边检验）

$$H_0: F(x) = F^*(x) \qquad \text{对所有 } x \in (-\infty, +\infty)$$
$$H_1: F(x) \neq F^*(x) \qquad \text{至少对某个 } x$$

如果 T 值超过了表 A13 中给出的双边检验的 $1-\alpha$ 分位数，则我们以显著水平 α 拒绝 H_0. 近似 p-值可以通过在表 A13 中插值得到，或者利用 2 倍的单边检验的 p-值

$$\text{单边}p\text{-值} = t\sum_{j=0}^{[n(1-t)]}\binom{n}{j}\left(1-t-\frac{j}{n}\right)^{n-j}\left(t+\frac{j}{n}\right)^{j-1} \qquad (7)$$

这里的 t 是检验统计量的观测值，且 $[n(1-t)]$ 是小于等于 $n(1-t)$ 的最大整数.

B.（单边检验）

$$H_0: F(x) \geqslant F^*(x) \qquad \text{对所有 } x \in (-\infty, +\infty)$$
$$H_1: F(x) < F^*(x) \qquad \text{至少对某个 } x$$

如果 T^+ 值超过了表 A13 中给出的单边检验的 $1-\alpha$ 分位数，则我们以显著水平 α 拒绝 H_0. 近似 p-值可以通过在表 A13 中插值得到. 精确的 p-值可以通过(7)式算出.

C.（单边检验）

$$H_0: F(x) \leqslant F^*(x) \qquad \text{对所有 } x \in (-\infty, +\infty)$$
$$H_1: F(x) > F^*(x) \qquad \text{至少对某个 } x$$

如果 T^- 值超过了表 A13 中给出的单边检验的 $1-\alpha$ 分位数，则我们以显著水平 α 拒绝 H_0. 近似 p-值可以通过在表 A13 中插值得到. 精确的 p-值可以通过（7）式算出.

432 **计算机辅助**　用软件 *S-Plus* 和 *StatXact* 计算 Kolmogorov 拟合优度检验. ◀

例 6.1.1

一组容量为 10 的样本如下：$X_1=0.621, X_2=0.503, X_3=0.203, X_4=0.477$, $X_5=0.710$, $X_6=0.581$, $X_7=0.329$, $X_8=0.480$, $X_9=0.554, X_{10}=0.382$. 零假设是样本服从分布函数为均匀分布函数，它的图像如图 6-3 所示. 假设的均匀分布函数的数学表达式为：

$$F^*(x) = \begin{cases} 0 & \text{若 } x < 0 \\ x & \text{若 } 0 \leqslant x < 1 \\ 1 & \text{若 } 1 \leqslant x \end{cases} \qquad (8)$$

形式上，假设检验给出如下：

$$H_0: F(x) = F^*(x) \qquad \text{对所有 } x \in (-\infty, +\infty)$$
$$H_1: F(x) \neq F^*(x) \qquad \text{至少对某个 } x$$

这里，$F(x)$ 是 X_i 的未知分布函数，$F^*(x)$ 由（8）式给出.

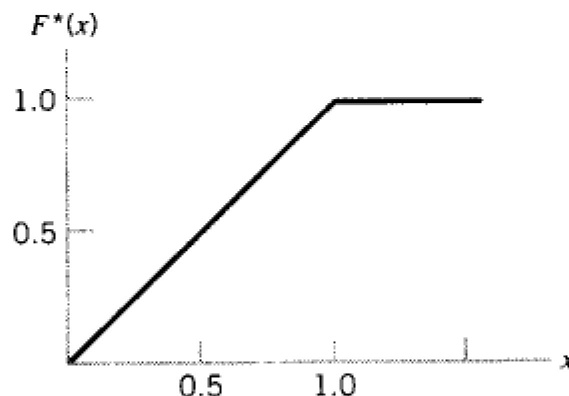


图 6-3　假设的分布函数

用双边 Kolmogorov 拟合优度检验. 水平为 $\alpha=0.05$ 的临界域对应着 T 值大于 0.95 分位数 0.409（由表 A20 当 $n=10$ 获得）. 由画出的经验分布函数 $S(x)$ 图，见图 6-4，T 的值在 $S(x)$ 的最高处（即 $x=0.710$ 处）取得，$S(x)$ 与 $F^*(x)$ 的最大垂直距离为 0.290，其中，$S(0.710)=1.000$ 和 $F^*(0.710)=0.710$. 也就是

$$T=\sup_x |F^*(x)-S(x)|$$
$$=|F^*(0.710)-S(0.710)|=0.290$$

因为 $T=0.290$，它小于 0.409，所以接受零假设. 从表 A13 可看出，p-值大于 0.20. 433

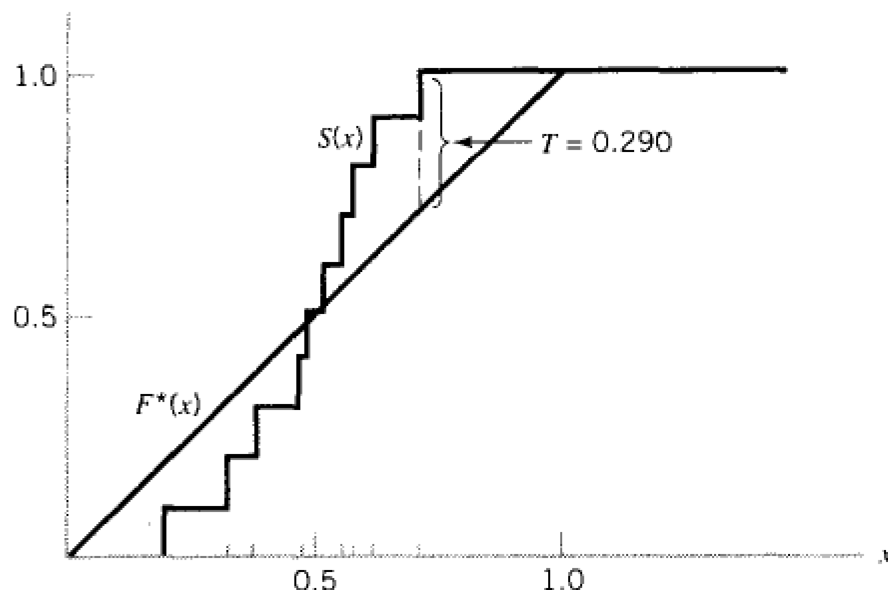


图 6-4 带有 T 值的 $F^*(x)$ 与 $S(x)$ 的图像

如果我们想检验零假设

$$H_0: F(x) \geqslant F^*(x) \quad 对所有的 x$$

对单边备择假设

$$H_1: F(x) < F^*(x) \qquad 对于某个 x$$

对此，我们用检验统计量 T^+，判决法则为：对于单边检验，如果 T^+ 超过 0.95 分位数，0.369（由表 A13 当 $n=10$ 获得），则我们以 $\alpha=0.05$ 的置信水平拒绝 H_0. 在这种情况下，可以计算 T^+ 的值，它在 $S(x)$ 左边的第 2 个跳跃点处取得.

$$T^+=\sup_x [F^*(x)-S(x)]=F^*(0.3289)-S(0.3289)$$
$$=0.3289-0.100=0.2289$$

更确切地说，$T^+=0.228999\cdots$，大约为 0.229. 最后的结果是相同的，p-值 >0.10.

对另一个方向的单边假设检验，结果为

$$T^-=\sup_x [S(x)-F^*(x)]=S(0.710)-F^*(0.710)$$
$$=1.000-0.710=0.290$$

p-值 >0.10.

对这种情形，双边检验是合适的检验，这里用单边检验，只是为了说明如何来估算检验统计量. 一般来说，双边检验统计量 T 总等于单边检验统计量 T^+ 和 T^- 中较大的一个. 434

双边检验中更精确的 p-值，可以通过（7）式中给出的 p-值的 2 倍来计算.

$$
\begin{aligned}
p\text{-值} &\doteq 2(0.29)\sum_{j=0}^{7}\binom{10}{j}\left(0.71-\frac{j}{10}\right)^{10-j}\left(0.29+\frac{j}{n}\right)^{j-1} \\
&= 2(0.29)(0.112+0.117+0.101+0.081+0.061+0.040+0.017+0.000) \\
&= 2(0.29)(0.530) \\
&= 0.307
\end{aligned}
$$

■

$F^*(x)$ 为离散时，一种计算精确 p-值的方法

如果所假设的分布$F^*(x)$为离散的，且从表 A13 中得到的 p-值不令人满意，那么精确的 p-值可以通过一个特殊的检验统计量观测值获得. 若样本量少于等于 5，计算过程可以用手算来完成. 若样本量较大，我们推荐用计算机程序，如 *StatXact* 来计算，其方法由下面给出. 对于离散型分布，p-值通常是由表 A13 所得到近似 p-值的 1/3. 下面的每一节对应于前面给出的 3 种假设检验 A,B 和 C.

A.（双边检验）　设 t 为检验统计量 T 的观测值. 接下来的 B 和 C 部分，用 t 代替 t^+ 和 t^- 也一样计算 $P(T^+\geqslant t)$ 和 $P(T^-\geqslant t)$. 那么

$$P(T\geqslant t)\doteq P(T^+\geqslant t)+P(T^-\geqslant t) \tag{9}$$

在大部分情况下，它是非常接近于真实 p-值的一个近似，除非 t 值很小时，它的值可能比真实的 p-值大.

B.（单边检验）　记 t^+ 为 T^+ 的观测值.

第 1 步：通过直接在$F^*(x)$的图像上画一条纵坐标为 $1-t^+-j/n$ 的水平线来计算概率 $f_j(0\leqslant j<n(1-t^+))$，则有 $f_j=1-t^+-j/n$，除非这条水平线与$F^*(x)$的一个

435 跳跃相交，在这种相交的情况下，f_j 等于相交处$F^*(x)$阶梯底部的高度值.

第 2 步：通过递归关系 $e_0=1$，且

$$e_k=1-\sum_{j=0}^{k-1}\binom{k}{j}f_j^{k-j}e_j \qquad k\geqslant 1 \tag{10}$$

来计算常数 $e_1,e_2,\cdots$，对于所有 k，使得第 1 步中的 $f_k>0$. 且注意，这些常数有如下形式：

$$
\begin{aligned}
e_0 &= 1 \\
e_1 &= 1-f_0 \\
e_2 &= 1-f_0^2-2f_1e_1 \\
e_3 &= 1-f_0^3-3f_1^2e_1-3f_2e_2 \\
e_4 &= 1-f_0^4-4f_1^3e_1-6f_2^2e_2-4f_3e_3 \\
e_5 &= 1-f_0^5-5f_1^4e_1-10f_2^3e_2-10f_3^2e_3-5f_4e_4
\end{aligned}
$$

等.

第 3 步：计算精确的单边 p-值

$$P(T^+\geqslant t^+)=\sum_{j=0}^{[n(1-t^+)]}\binom{n}{j}f_j^{n-j}e_j \tag{11}$$

其中，f_j，e_j 为第 1，第 2 步中所得到的.

C.（单边检验） 记 t^- 为 T^- 的观测值.

第 1 步：计算概率 $c_j(0 \leqslant j < n(1-t^-))$ 如下. 在$F^*(x)$的图像上，画一条纵坐标为 $t^- + j/n$ 的水平线，则 $c_j = 1 - t^- - j/n$. 除非这条水平线与$F^*(x)$的一个跳跃相交，在这种相交的情况下，c_j 等于 1 减去相交处$F^*(x)$阶梯顶部的高度值.

第 2 步：通过递归关系 $b_0 = 1$，且

$$b_k = 1 - \sum_{j=0}^{k-1} \binom{k}{j} c_j^{k-j} b_j \qquad k \geqslant 1 \tag{12}$$

来计算常数 $b_1, b_2, \cdots$，对于所有 k，使得第 1 步中的 $c_k > 0$. 这些常数与 B 部分的 e_k 有相同的形式，只需用 c_j 替代 f_j 即可. 436

第 3 步：计算精确的单边 p-值

$$P(T^- \geqslant t^-) = \sum_{j=0}^{[n(1-t^-)]} \binom{n}{j} c_j^{n-j} b_j \tag{13}$$

其中 c_j，b_j 为第 1，第 2 步所得到的.

下面的例子将说明，当$F^*(x)$为离散时，如何用这种方法来计算精确的p-值.

例 6.1.2

令$F^*(x)$为离散的均匀分布，在 $x = 1, 2, 3, 4, 5$ 有相同的概率 1/5. 设有一组样本容量为 10 的抽自某个总体的随机样本（排序）如下：1，1，1，2，2，2，3，3，3，3. 而零假设为$F^*(x)$就是这个总体的分布函数，而$F^*(x)$与$S(x)$的最大距离出现在 $x = 3$ 处（如图 6-5），那么双边 Kolmogorov 检验统计量为

$$T = \sup_x |F^*(x) - S(x)| = 0.4 = t \tag{14}$$

为了计算 $t = 0.4$ 时的 p-值，首先我们计算 $P(T^+ \geqslant 0.4)$ 的概率.

第 1 步：因为 $n(1-t) = 10(0.6) = 6$，所以需用计算 f_0 到 f_5. 纵坐标为$1 - t = 0.6$ 的水平线与$F^*(x)$在跳跃点 $x = 3$ 处相交，所以 f_0 等于水平线的纵坐标，$f_0 = 0.6$. 对于 $j = 1$，水平线 $1 - t - 1/10 = 0.5$ 与$F^*(x)$的一个跳跃相交，所以 f_1 等于相交处$F^*(x)$阶梯底部的高度值，即：$f_1 = 0.4$. 同样地，$f_2 = 0.4$，$f_3 = 0.2$，$f_4 = 0.2$，$f_5 = 0$.

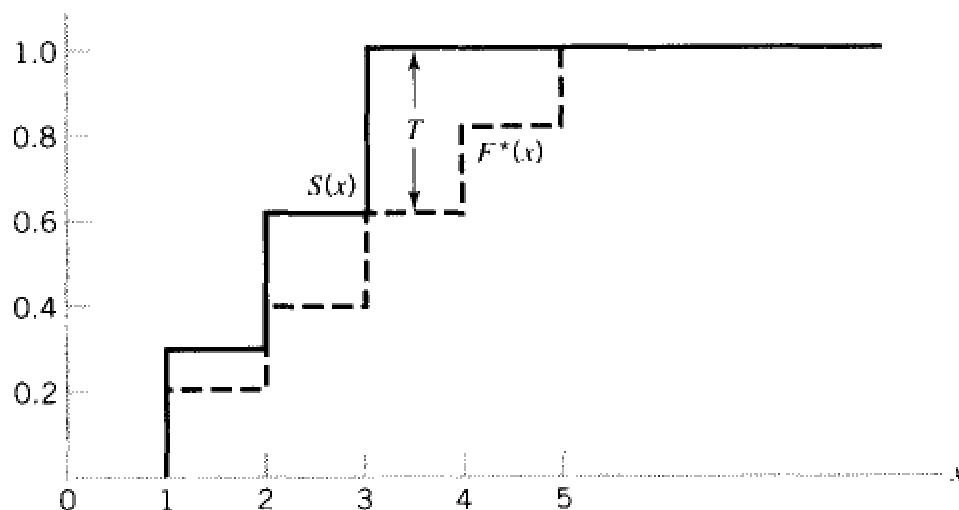


图 6-5 带有 T 值的$F^*(x)$与 $S(x)$的图像 437

第 2 步：从（10）式可递归计算常数 e_0 到 e_4：

$$e_0 = 1$$

$$e_1 = 1 - 0.6 = 0.4$$

$$e_2 = 1 - (0.6)^2 - 2(0.4)(0.4) = 0.32$$

$$e_3 = 1 - (0.6)^3 - 3(0.4)^2(0.4) - 3(0.4)(0.32) = 0.208$$

$$e_4 = 1 - (0.6)^4 - 4(0.4)^3(0.4) - 6(0.4)^2(0.32) - 4(0.2)(0.208) = 0.2944$$

第 3 步：利用（11）式，计算单边 p-值 $P(T^+ \geqslant t)$

$$P(T^+ \geqslant t) = f_0^{10} + \binom{10}{1} f_1^9 e_1 + \binom{10}{2} f_2^8 e_2 + \binom{10}{3} f_3^7 e_3 + \binom{10}{4} f_4^6 e_4 = 0.021$$

因为 $F^*(x)$ 是对称的，计算另一边的单边 p-值 $P(T^- \geqslant 0.4)$ 与前面的过程相同．所以 $P(T^- \geqslant 0.4) = 0.21$，且双边 p-值近似为：

$$P(T \geqslant 0.4) \doteq 2(0.021) = 0.042$$

有趣的是，注意到这个 p-值表明在 $\alpha = 0.05$ 的水平下，正确的判决是拒绝零假设，而用表 A13 将导致在同样的水平 α 下，错误地接受 $F^*(x)$ 为正确的分布函数．■

评注

Kolmogorov 双边检验一个最有用的特征是，它的 $1-\alpha$ 分位数 $w_{1-\alpha}$ 可以用来为真实未知的分布函数构造一个置信界．回忆一下，我们在为某个未知参数寻找置信区间时，首先抽出一组随机样本，然后从这组随机样本中计算一个上界值 U 和一个下界值 L，使得它以 $1-\alpha$ 的概率包含这个未知参数，并把 $1-\alpha$ 称为置信系数．这里一个方便的做法是，如果我们能够对完全未知的分布函数做同样的事情来获得一个“置信界”，使得完全未知的分布函数以 $1-\alpha$ 的概率落在这一置信界内．然后我们可以从某个完全未知的分布总体中抽取样本，并且能为其图像设定一个界，说明这个未知的分布函数以 $1-\alpha$ 的概率正确地落在这个界内．

▶总体分布函数的置信界

数据　数据包括容量为 n 的，且来自于某个总体的随机样本 $X_1, X_2, \cdots, X_n$，总体的未
438 知分布函数记为 $F(x)$．

假定条件

1. 样本是随机样本．

2. 为了置信系数的精确，随机变量应当是连续的．如果随机变量是离散的，那么这个置信界是保守的，也就是说未知真实的置信系数大于我们所给的．

方法　画一个随机样本的经验分布函数 $S(x)$ 图．为了构造置信系数为 $1-\alpha$ 的“置信界”，我们可以从表 A13 中查到双边检验（如果想要构造双边置信界），及合适的样本容量为 n 时 Kolmogorov 检验统计量的 $1-\alpha$ 分位数．用 $w_{1-\alpha}$ 记此分位数，在 $S(x)$ 的上方距离为 $w_{1-\alpha}$ 处画一图像，称为 $U(x)$，在 $S(x)$ 的下方距离为 $w_{1-\alpha}$ 处画第

2 个图像，称为 $L(x)$. 于是 $U(x)$ 与 $L(x)$ 就分别形成了上、下界，使得完全未知的分布函数 $F(x)$ 以 $1-\alpha$ 的置信水平落在这个界内.

注意，$U(x)$ 的图像不能超过 1.0，即使 $S(x)+w_{1-\alpha}$ 可能超过 1.0，因为我们知道分布函数的值不可能超过 1.0. 同样的原因，$L(x)$ 的值不能低于横轴的值. $U(x)$ 与 $L(x)$ 的数学表达式如下

$$\begin{aligned} U(x) &= S(x) + w_{1-\alpha} \quad 若 \quad S(x) + w_{1-\alpha} \leqslant 1 \\ U(x) &= 1.0 \quad 若 \quad S(x) + w_{1-\alpha} > 1 \end{aligned} \tag{15}$$

$$\begin{aligned} L(x) &= S(x) - w_{1-\alpha} \quad 若 \quad S(x) - w_{1-\alpha} \geqslant 0 \\ L(x) &= 0 \quad 若 \quad S(x) - w_{1-\alpha} < 0 \end{aligned} \tag{16}$$

用概率的语言表达为：

$$P[L(x) \leqslant F(x) \leqslant U(x), 对所有的 x] \geqslant 1-\alpha \tag{17}$$

其中，最后一个不等号只有当随机变量为离散时成立.

例 6.1.3

假设我们要为未知分布函数构造一个 90% 的置信界，一组容量为 20 的随机样本抽自这个总体，结果按从大到小排列如下：

16.7	17.4	18.1	18.2	18.8	19.3	22.4	22.4	24.0	24.7
25.7	27.0	35.1	35.8	36.5	37.6	39.8	42.1	43.2	46.2

439

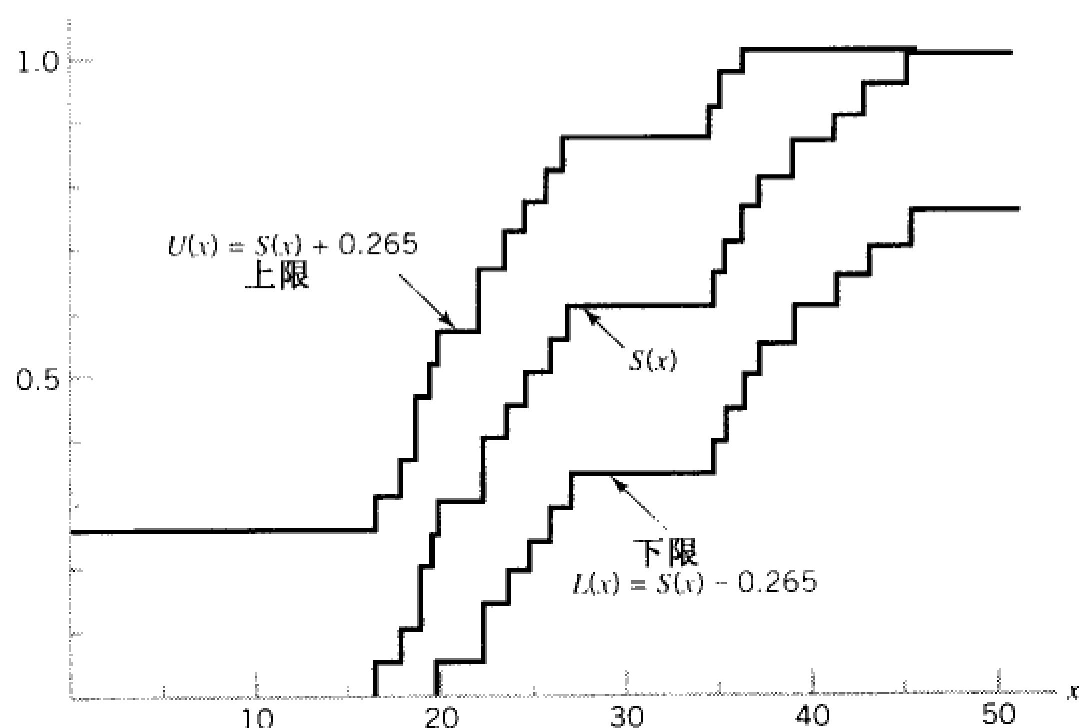


图 6-6 $F(x)$ 的置信界

从表 A13 可知，当 $n=20$ 时，0.90 分位数 $w_{0.90}=0.265$. 这个置信界就是 $S(x) \pm 0.265$，只要界位于 0 与 1 之间. 图 6-6 画出了 $S(x)$，$U(x)$ 和 $L(x)$. 结论"$F(x)$ 完全界于 $U(x)$ 与 $L(x)$ 之间"成立的概率为 0.90. ■

Kolmogorov 统计量的分布的推导很复杂，这里我们没有写出. 有兴趣的读者可以参考以下有关这方面的一些基本论文.

双边检验统计量 T 的渐近分布是由 Kolmogorov (1933) 发现的，并由 Smirnov

(1948)制成表格形式. 单边检验统计量 T^+ 和 T^- 的渐近分布是 Smirnov(1939)获得的. 对于有限（小）样本的检验统计量的精确分布, Wald 和 Wolfowitz(1939)进行过研究, 并由 Massey(1950a)制成表格. T^- 的有限样本分布函数由 Birnbaum 和 Tingey (1951)推导出, 并把从他们的精确分布获得的精确分位数与 Smirnov(1939,1948)给出的渐近分位数做了比较, 发现用渐近的分位数将导致保守的检验.

只要 $F(x)$ 与 $F^*(x)$ 有差异, 双边 Kolmogorov 检验就有一致拒绝的良好性质, 其中 $F(x)$ 为真实的分布, $F^*(x)$ 为假设分布. 然而, 对于有限样本来说, 它是有偏的(Massey,1950b). Massey(1950b)还给出了双边检验功效的一个下界, 而在某些特定的备择假设下, Birnbaum(1953)给出了功效的下确界, 在另一类不同的备择假设下, Lee(1966)给出了功效的另一个下确界.

Van der Waerden(1953), Suzuki(1968), Shapiro, Wilk 和 Chen(1968)以及 Knott (1970) 对功效作了其他的比较. 关于 Kolmogorov 检验及其类似的拟合优度检验问题

440 的文章可见 Finkelstein 和 Schafer(1971), Maag 和 Dicaire(1971), Carnal 和 Riedwyl (1972)以及 Stephens(1974). Barr 和 Davidson(1973), Pettitt 和 Stephens(1976) 在对截尾数据的情况下, 对检验提出了修改, 而 Barr 和 Shudde(1973)讨论了对圆周上观测数据检验的修改. Govindarajulu 和 Klotz(1973)对渐近分布作了注解. 关于估计与检验对称分布为主题的文章参见 Schuster 和 Narvarte(1973), Schuster(1973)以及 Sirnivasan 和 Godio(1974).

对于在离散型分布情形下的修正则由 Conover(1972)与 Coberly 和 Lewis(1973)独立提出的. 关于这种情形下更深入的分析, 可参见 Horn 和 Pyne(1976), Horn (1977), Bartels, Horn, Liebetrau 和 Harris(1977)与 Pettitt 和 Stephens(1977)的文章, 同样他们还制作了一些表格. Maag, Streit 和 Drouilly(1973)讨论了分组数据的拟合优度检验. Wood 和 Altavela(1978)建议对大样本采用模拟的方法.

另外一种拟合优度检验就是 Cramér-von Mises 检验, 由 Cramér(1928), Van Mises(1931) 与 Simrnov(1936)发展起来的. 尽管相对于 Kolmogorov 检验, 对于许多人来说它更有直观性, 但是这两种检验并没有太大的区别, 我们这里不做详细讨论. 有兴趣的读者, 可以查看 Anderson 和 Darling(1952)给出的 Cramér-von Mises 检验的渐近分布. Stephens 和 Maag(1968) 给出了有限样本的精确表格. 较早地对这种检验与 Kolmogorov 检验进行研究的有 Stephens(1964, 1965a), Tiku(1965), Suzuki (1967), Cronholm(1968) 与 Noé 和 Vanderwiele(1968). Walsh(1960,1963)讨论了这两种检验的离散（结）效应. Thompson(1966) 考查了 Cramér-vonMises 检验的偏差与功效. Gelzer 和 Pyke(1965), Quade(1965)以及 Abrahamson(1967) 研究了 Kolmogorov 检验的相对效率.

关于样本密度的拟合优度检验问题, Woodroofe(1967)提出并进行了讨论, Stephens(1967)对圆周上观测的情况进行了讨论, Riedwyl(1967) 对一般情况进行了讨论. Durbin(1968)介绍了一种不同类型的分布函数的置信区间.

习题

1. 从一个四年级班上随机抽出 5 名学生进行计时短跑，他们的时间（秒）分别为 6.3,4.2, 4.7, 6.0 和 5.7. 用画图或列表格的形式给出这个四年级班所有学生短跑时间分布函数的 90% 置信界.
2. 乡村的杂货店从近邻农家收取鸡蛋时，需要在灯光前检查鸡蛋是否新鲜. 现有 8 箱鸡蛋，每箱有 144 个. 经检查每箱中被拒收的鸡蛋个数为：4,0,2,0,2,0,2,0. 试用画图或列表格的形式给出所有收取鸡蛋的总体中被拒绝鸡蛋的个数分布函数的 95% 置信界.
3. 对于习题 1 中给出的数据，检验其是否服从 4 到 8 秒间的均匀分布. 注意，这个分布函数为 441

$$
\begin{aligned}
F^*(x) &= 0 && \text{对} \quad x < 4 \\
&= (x-4)/4 && \text{对} \quad 4 \leqslant x < 8 \\
&= 1 && \text{对} \quad 8 \leqslant x
\end{aligned}
$$

4. 以前的记录表明每箱中被拒绝鸡蛋的个数服从均值为 1.5 的 Possion 分布，对于习题 2 中的数据检验其是否来自这样的分布. 注意，均值为 1.5 的 Possion 分布有如下概率：$P(0)=0.223, P(1)=0.335, P(2)=0.251, P(3)=0.126, P(4)=0.047, P(5)=0.014, P(6)=0.004$.
5. 奥运会跳水比赛要用 10 次跳水来评定成绩，其 10 次成绩为：1.7,5.3,7.6,8.9,9.0,9.1, 9.3,9.6,9.9,9.9. 检验其是否服从如下的分布 $F(x)$，其中

$$
\begin{aligned}
F(x) &= 0 && \text{若} x < 0 \\
F(x) &= x^2/100 && \text{若} 0 \leqslant x \leqslant 10 \\
F(x) &= 1 && \text{若} 10 < x
\end{aligned}
$$

6. 通过上一年对数千辆汽车排放的氮氧化物的测量，发现它大约服从均值为 5.6，标准差为 1.2 的正态分布模型. 今年测量的 12 辆汽车的排放量为：

4.8, 6.2, 6.0, 5.9, 6.6, 5.5, 5.8, 5.9, 6.3, 6.6, 6.2, 5.0

问今年汽车排放量的分布模型是否与去年的相同?

思考题

证明(17)式中的置信界是有效的，即证明：如果 $w_{1-\alpha}$ 是 Kolmogorov 统计量的 $1-\alpha$ 分位数，则（17）式成立.

6.2 分布族的拟合优度检验

6.1 节中我们讲述的 Kolmogorov 拟合优度检验，是用来检查一个样本是否符合某个特定分布的一种很好的检验方法. Kolmogorov 检验只有当假设的分布完全已知的时候才适用，也就是说，假设的分布不包含需要从样本里估计的未知参数，否则，检验就变得保守了. 而 χ^2 拟合优度检验则比较灵活，允许分布中包含有需要从数据中待估的未知参数. 正如前面所描述的，用“最小 χ^2 估计”估计每一个参数，则自由度需要减

去一个. 然而，χ^2 检验要求数据必须分组，而这种分组又通常是随意的. 同时，检验
442 统计量的分布只是近似的，且有时χ^2 检验的功效也不高. 由于这些原因，我们想寻找别的拟合优度检验，特别是对一些经常需要检验的分布.

Kolmogorov 检验适当修改后也能适合许多包含待估参数的情况. 实际上，检验统计量并未改变，只是用到的临界值表有变化. 对于所有的分布，这些表不再是用同样一些表. 它随着不同的假设分布而变化. 这种检验仍然是非参数的，因为这个检验的有效性（α 水平）不依赖于关于总体分布未检验的具体假设条件，而检验的是总体的分布形式假设.

Kolmogorov 检验最先的修改用于检验复合正态分布的假设. 也就是说，零假设只是说明总体来自正态分布族，但是未指明正态分布的均值与方差，Lilliefors(1967) 最早提出了这种检验. 对这个检验有意思的一点是，为了获得检验统计量精确分布的真实分位数的精确估计，一个最早的方法就是采用计算机生成随机数的方法.

▶Lilliefors 正态性检验

数据 数据包含来自某个未知分布的 n 个随机样本 $X_1, X_2, \cdots, X_n$，未知分布的分布函数记为 $F(x)$. 计算样本均值

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

作为 μ 的一个估计；计算

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} \tag{2}$$

作为 σ 的一个估计. 然后计算“标准化”后的样本值 Z_i，它定义为

$$Z_i = \frac{X_i - \overline{X}}{s} \qquad i = 1, 2, \ldots, n \tag{3}$$

检验统计量是在原检验统计量中用 Z_i 替代原始随机样本而得到.

假定条件

443 1. 样本是随机样本.

检验统计量 通常情况下，检验统计量是双边 Kolmogorov 检验统计量，定义为 X_i 的经验分布函数与均值为 $\bar{X}$，标准差为 s 的正态分布的最大垂直距离，其中 $\bar{X}$，s 由(1),(2)式给定. 然而，下面计算检验统计量的方法稍微容易一些，且与我们上面所提到的方法是等价的.

画一个标准正态分布函数的图像，记为$F^*(x)$. 实际上，我们只需要$F^*(x)$在观测点 Z 处的值，这时表 A1 可能起到作用. 同样我们画一个“标准化”样本 Z_i 的经验分布函数的图像，Z_i 的定义见(3)式，记 $S(x)$ 为其经验分布函数，与$F^*(x)$用同一个坐标系，找出$F^*(x)$与 $S(x)$的最大垂直距离，这个距离就是检验统计量. 也就是说，Lilliefors 检验统计量 T_1 定义如下：

$$T_1 = \sup_x |F^*(x) - S(x)| \tag{4}$$

T_1 与 Kolmogorov 检验统计量的不同之处在于(4)式中用的经验分布函数是由标准化的样本得来的，而 Kolmogorov 检验则是基于原始观测样本.

零分布 在计算机上产生几千个伪随机数来近似获得零分布，并且通过数千个检验统计量的值作出的经验分布函数来估计分位数. 表 A14 中给出了这些分位数的估计，精确的分位数及精确的零分布的数学形式仍然未知.

假设

H_0: 样本来自于一个未知均值和标准差的正态分布总体
H_1: X_i的分布函数不是正态分布

如果 T_1 超过了表 A14 中给出的 $1-\alpha$ 分位数，则我们以近似置信水平 α 拒绝 H_0. p-值可以通过表 A14 中的分位数近似求得.

计算机辅助 Lilliefors 正态性检验可以在 *Minitab*, *S-Plus*, 及 *StatXact* 中进行计算. —◀

例 6.2.1

在 4.5.3 节中，我们用实例来解释 χ^2 正态性检验，现在我们用同样的数据来说明 Lilliefors 检验. 从电话本里随机抽取 50 个 2 位数，显然随机变量的样本是离散的， 444
尽管如此，我们依然用它作正态性检验. 只要我们注意到接受正态性的零假设并不意味着随机变量就是正态的，因而是连续的，而仅仅表明正态分布函数与它实际分布函数的差别不显著，以至于不能检测出来.

将 X_i 从小到大排列，并且从（1）和（2）式中分别减去 $\bar{X}=55.04$，除以 $s=19.00$，这时 X_i 转化为 Z_i，如下表：

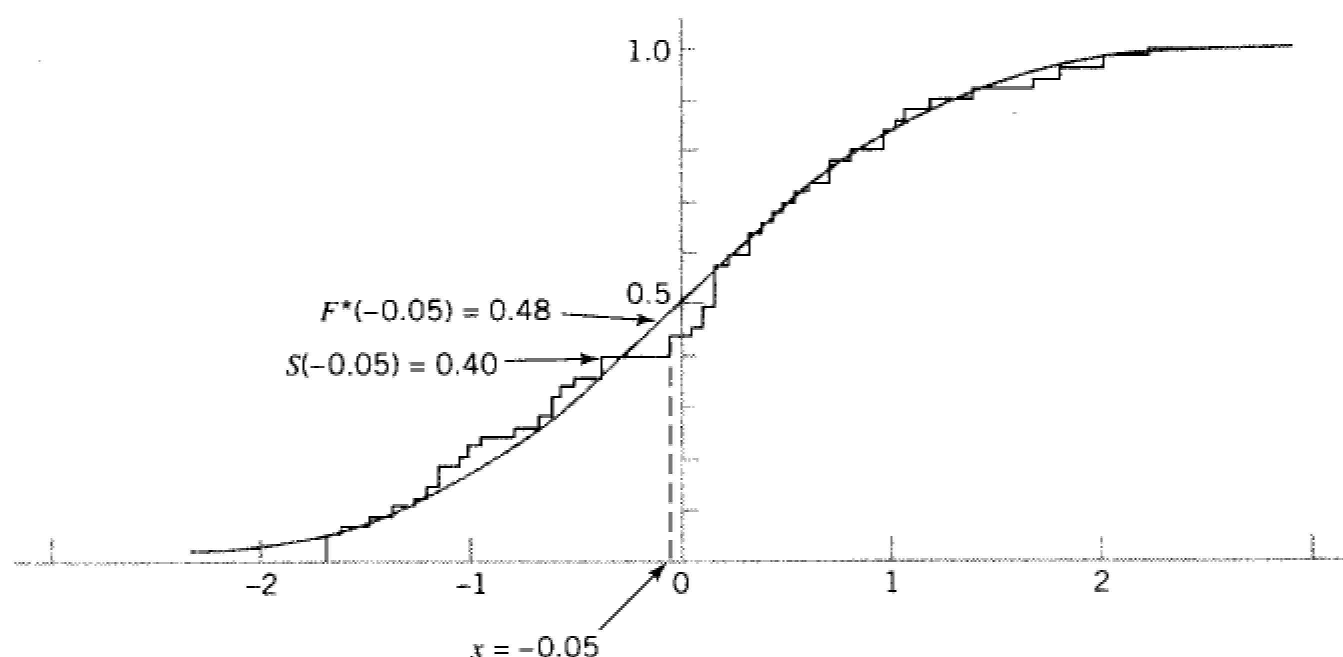
X_i	Z_i	X_i	Z_i	X_i	Z_i	X_i	Z_i	X_i	Z_i
23	−1.69	36	−1.00	54	−0.05	61	0.31	73	0.95
23	−1.69	37	−0.95	54	−0.05	61	0.31	73	0.95
24	−1.63	40	−0.79	56	0.05	62	0.37	74	1.00
27	−1.48	42	−0.69	57	0.10	63	0.42	75	1.05
29	−1.37	43	−0.63	57	0.10	64	0.47	77	1.16
31	−1.27	43	−0.63	58	0.16	65	0.52	81	1.37
32	−1.21	44	−0.58	58	0.16	66	0.58	87	1.68
33	−1.16	45	−0.53	58	0.16	68	0.68	89	1.79
33	−1.16	48	−0.37	58	0.16	68	0.68	93	2.00
35	−1.05	48	−0.37	59	0.21	70	0.79	97	2.21

用 Lilliefors 检验统计量进行正态性的零假设检验，

$$T_1 = \sup_x |F^*(x) - S(x)|$$

其中，$F^*(x)$是标准正态分布函数，$S(x)$是 Z_i 的经验分布函数. 如图 6-7，画出了 $F^*(x)$与 $S(x)$的图像. 在图 6-7 中我们可以看到，$F^*(x)$与 $S(x)$之间的最大垂直距离出现在 $x=-0.05$ 的左端，此时$F^*(x)=0.48$，$S(x)=0.40$，故 $T=0.08$. 但是两条曲线的垂直距离等于0.08还出现在其他点上，如$x=+0.05$与$x=0.10$. 但是

没有点使两条曲线的距离超过 0.08.



445 图 6-7 $F^*(x)$ 与 $S(x)$ 的图像及其它们之间的最大距离

Lilliefors 正态性检验的检验法则是：如果 T_1 超过 0.95 分位数，则我们以 $\alpha =$ 0.05 水平拒绝 H_0，在表 A14 中给出的 0.95 分位数为

$$w_{0.95} = \frac{0.886}{\sqrt{n}} = \frac{0.886}{\sqrt{50}} = 0.125$$

由于 $T_1 = 0.08$，且小于 0.125，所以接受零假设. 实际上，在 $\alpha = 0.20$ 的水平上，也可以接受零假设，因为 0.8 分位数等于 0.104. 又由于表 A14 中未给出更小的分位数，所以我们断言 p-值一定是某个大于 0.20 的数，这与我们在 χ^2 检验中得出的结论相同.

接受零假设并不意味着母体是正态的，但它表明正态分布不失为其真实未知分布的一个合理的近似. 因此对于这组数据，不管是非参数的方法，还是参数的方法都表明正态性母体假设是合适的. ■

□**理论** 我们讨论 Lilliefors 检验的一个主要原因是，说明如何得到表 A14 中的分位数. 寻求 T_1 的分布使得 Kolmogorov 检验可以用来检验均值方差未知的复合正态性假设是很困难的问题，它没有解析表达式. 因此，Lilliefors 用计算机和随机数找到了近似解. 下面描述的同样方法可以为大多数的统计推断问题找到近似解.

回忆一下，为了作统计假设检验，首先必须找到一个合理的检验统计量，作为表明零假设正确与否的敏感标识，T_1 就满足这种要求. 其次，还必须选定某个特定的区域作为对应的临界域，也就是若 H_0 是正确的，临界域中的值不太可能出现；但若 H_0 是错误的，那么临界域中的值出现的可能性较大，较大的 T_1 值满足这个要求. 然后，寻求 α（即当 H_0 为真时，点出现在临界域中的概率）是比较困难的. 为了找 α 值，Lilliefors 在高速计算机上产生随机正态离差（5.10 节中提到），它是一些看似像独立标准正态随机变量观测值的随机数.

把这些随机数分组，样本容量为 n 分为一组，n 可以是随意的. 比如，$n = 8$，从

一个标准正态分布中产生一个容量为 8 的样本，由于是计算机产生，所以认为 H_0 是正确的. 计算样本均值 $\bar{X}$，并把它减去，结果再除以由（2）式算出的 s，最后得到 Z_i 的值. 将基于这些 Z_i 的经验分布函数与标准正态分布函数比较，并记录下它们之间最大垂直距离 T_1. 对于另外一组 8 个数据重复上述过程，得到的另一个 T_1 的值. 最后，在 H_0 为真的情况下，得到超过 1000 组（每组样本容量为 8）的样本以及超过 1000 个的 T_1 值. 基于这 1000 个或更多的 T_1 值的经验分布函数，可以作为 T_1 未知实际分布的一个近似. 从这个经验分布函数中，我们得到如表 A14 中 $n=8$ 时所给的分位数. 这样，对于某些 α，就能得到近似水平为 α 的临界域. 446

对于 $n=4$ 到 $n=30$，我们可以重复上述过程. 为了得到 n 大于 30 时 Lilliefors 所提出的近似，譬如 $n=40$，分位数还是由上面的方法给出，并乘以 $\sqrt{40}$后得到表中的数据. 这个方法是基于一个未被证明的猜想（在当时）：T_1 趋于它的渐近分布的速度与 Klomogorov 统计量趋于其极限分布的速度一致，即$\sqrt{n}$的速度.

后来，为了获得更准确的分位数估计，Mason 和 Bell(1986)用 $n=20\,000$ 进行了模拟，结果列在表 A14 中.

Lilliefors(1967)还比较了他的检验功效与 χ^2 检验的功效，发现在一些非正态的情况下，他的检验功效更高. □

Srinivasan 和 Wharton(1973)得到了参数形式下的正态分布的一个参数置信界. 其他相关的论文有 Srinivasan(1972)及 Dyer(1974). Teichroew(1965) 给出了模拟及其模拟的时间的一般讨论.

Lilliefors 对 Kolmogorov 检验的另一个改进是在 1969 年，它是用来检验母分布函数是否为指数分布，即 $F(x)=1-e^{-x/t}(x>0)$，其中 t 是未知的需要用数据来估计的参数（$e=2.718\cdots$是一个熟知的常数）. 尽管 Lilliefors 利用上述的随机数模拟得到了近似的临界值，但是其检验统计量的精确分布随后由 Durbin(1975)与 Margolin 和 Maurer(1976)年推导出. 由于他们使用的方法超出了本书的范围，因此在这里就不再做介绍了.

一般理论表明，指数分布通常是用来描述一系列随机事件随时间发生时，中间间隔的时间长度，因此检验指数分布实际上主要用来检验随机性. 这个检验的另一个应用是，在产品的寿命内，检测产品的失效率是否为常数，这也是指数分布的一个特性. 447

▶指数分布的 Lilliefors 检验

数据 数据包含来自某未知分布的 n 个随机样本 $X,X_2,\cdots,X_n$，未知分布的分布函数记为 $F(x)$. 计算样本均值作为未知参数的一个估计，对每个 X_i，如下定义 Z_i：

$$Z_i = X_i/\overline{X} \tag{5}$$

它将用于检验统计量的计算中.

假定条件

1. 样本是随机样本.

检验统计量 先画一个基于 Z_1，…，Z_n 的经验分布函数图像 $S(x)$，并在同一个图上画出函数 $F^*(x)=1-e^{-x}(x>0)$ 的图像. 实际上，只需要确定 n 个点上的值便可，即 $x=Z_1, x=Z_2$，等等. 可以用查表或计算器计算 e^{-x}. 两个函数间的最大垂直距离

$$T_2=\sup_x |F^*(x)-S(X)| \tag{6}$$

就是检验统计量.

尽管这只是一个双边检验，但是 Durbin(1975) 给出了单边检验及其计算临界值所用的表.

零分布 尽管刚开始零分布的分位数由计算机模拟产生随机数来估计，随后的研究成功地给出了其精确的分布. 精确分位数列在表 A15 中.

假设检验

H_0: **随机样本服从指数分布**

$$F(x)=\begin{cases}1-e^{-x/t}, & x>0\\ 0, & x<0\end{cases} \tag{7}$$

其中 t 为未知参数

448 H_1: **X的分布不是指数分布**

检验法则为：如果 T_2 超过表 A15 中的 $1-\alpha$ 分位数，则我们以显著性水平 α 拒绝 H_0. 近似 p-值由在表 A15 中插值得到.

计算机辅助 *S-Plus* 可以做 Lilliefors 指数分布检验. ◀

例 6.2.2

一般认为长途电话通过电话总机接通的过程是一个随机过程，其间打进电话的时间间隔服从指数分布. 某个星期一下午 1:00 以后最先打进的 10 个电话发生在 1:06, 1:08, 1:16、1:22, 1:23, 1:34, 1:44, 1:47, 1:51, 1:57. 第一个电话时间间隔，即从 1:00 到 1:06 为 6（分钟），后面的时间间隔（分钟）分别为：2, 8, 6, 1, 11, 10, 3, 4, 6. 算得其样本均值为 $\bar{X}=5.7$，这样就得到 Z_i, $1-e^{z_i}$ 以及 $S(x)$ 和 $F^*(x)$ 之间的距离（在 $S(x)$ 的每个跳跃点两端）在下表中给出. 为方便起见，我们把数据 X 按从小到大排列.

i	X_i	$Z_i=X_i/\bar{X}$	$1-e^{-z_i}$	$i/10-1+e^{-z_i}$	$1-e^{-z_i}-(i-1)/10$
1	1	0.1754	0.1609	−0.0609	0.1609
2	2	0.3508	0.2959	−0.0959	0.1959
3	3	0.5263	0.4092	−0.1092	0.2092
4	4	0.7018	0.5043	−0.1043	0.2043
5	6	1.0526	0.6510	−0.1510	0.2510[b]
6	6	1.0526	0.6510	−0.0510	0.1510
7	6	1.0526	0.6510	0.0490	0.0510
8	8	1.4035	0.7543	0.0457	0.0543
9	10	1.7544	0.8270	0.0730	0.0270
10	11	1.9298	0.8548	0.1452[a]	−0.0452

[a] $S(x)-F^*(x)$ 的最大距离.

[b] $F^*(x)-S(x)$ 的最大距离.

$S(x)$与$F^*(x)$的最大垂直距离为 0.2510. 在 $\alpha=0.05$ 的水平下，如果 T_2 超过 0.3244（从表 A15 中获得，$n=10$，$1-\alpha=0.95$），我们将拒绝指数分布作为零假设. 因为 $T_2=0.2510$，所以接受零假设. 通过在表 A15 中插值，可得 p-值为 0.25. 长途电话等待时间因此可认为是一个随机过程. ■

Lilliefors(1973)与 Schneider 和 Clickner(1976)把 Kolmogorov 检验推广到了带有待估参数的 Gamma 分布情形. 同样类型的 Cramér - von Mises 检验由 Pettitt 于 1978 年给出. Green 和 Hegazy(1976)讨论了类似的其他检验.

在这一节的最后，我们介绍一个著名的正态性拟合优度检验，它在很多情况下（如果愿意的话）可用来代替 Lilliefors 检验. 一些经验的研究表明，在检验复合正态 449
性假设时，许多情形下这种检验比其他检验，包括 Lilliefors 检验和χ^2 检验，有更高的功效（见 Shapiro, Wilk 和 Chen 1968; La Brecque 1977）. 尽管这种检验不是 Kolmogorov 型检验，但由于它的作用，我们仍在此讲述.

▶Shapiro-Wilk 正态性检验

数据 数据包含来自某未知分布的 n 个随机样本 $X_1, X_2, \cdots, X_n$，未知分布的分布函数记得为 $F(x)$.

假设条件

1. 样本是随机样本.

检验统计量 首先计算检验统计量的分母 D：

$$D=\sum_{i=1}^{n}(X_i-\overline{X})^2 \tag{8}$$

其中，$\bar{X}$ 是样本均值. 然后将样本从小到大排序

$$X^{(1)} \leqslant X^{(2)} \leqslant \cdots \leqslant X^{(n)}$$

且记 $X^{(i)}$ 为第 i 个次序统计量，对 n 个样本观测值，从表 A16 中得到系数$a_1, \cdots, a_k$，其中 k 大约为 $n/2$.

检验统计量 T_3 如下给出：

$$T_3=\frac{1}{D}\left[\sum_{i=1}^{k} a_i\left(X^{(n-i+1)}-X^{(i)}\right)\right]^2 \tag{9}$$

注意，这个检验统计量常记为 W，这个检验一般称为 W 检验.

零分布 检验统计量 T_3 是相关系数的平方，这里计算的相关系数是次序统计量 $X^{(i)}$ 与得分 a_i 之间的 Pearson 相关系数，其中 a_i 代表的是，当总体为正态分布时，次序统计量会有什么样的得分值. 因此如果 T_3 接近 1.0，那么样本近似于服从正态分布. 如果 T_3 太小，也就是远小于 1.0，那么样本非正态. 表 A17 中给出了 T_3 的分位数.

假设

H_0: $F(x)$是均值和标准差未知的正态分布 450

H_1: $F(x)$ 不是正态分布

如果 T_3 超过了表 A17 中给出的 α 分位数，则我们以 α 水平拒绝 H_0. 如果我们对 T_3 的一个观测值想要更精确的 p-值，由表 A18，我们可以把 T_3 转化为近似正态的随机变量，然后通过表 A1 与正态变量分位数比较而获得近似的p-值.

计算机辅助　*Minitab*, *SAS*, *StatXact* 可以做 Shapiro-Wilk 检验. ◄

评注

尽管已知的表格只允许作 $n \leq 50$ 的 Shapiro-Wilk 检验，但是 D'Agostino（1971）提出了一个可以用于 $n > 50$ 时的检验，且 Shapiro 和 Francia(1972)建议在 $n > 50$ 时做一种与 Shapiro- Wilk 检验类似的近似检验.

例 6.2.3

在例 4.5.3 中，我们给出了抽自电话本的 50 个 2 位数. χ^2 检验表明接受正态性假设，并得到 p-值为 0.25. 在例 6.2.1 中，Lilliefors 检验表明接受同样的假设，且 p-值大于 0.20. 同样对于这组数据，我们作 Shapiro-Wilk 检验.

从表 A16 中得到的系数 a_i 及次序统计量之差 $X^{(n-i+1)} - X^{(i)}$，如下所示：

	i	a_i	$X^{(n-i+1)} - X^{(i)}$	i	a_i	$X^{(n-i+1)} - X^{(i)}$
	1	0.3751	97–23	14	0.0846	66–42
	2	0.2574	93–23	15	0.0764	65–43
	3	0.2260	89–24	16	0.0685	64–43
	4	0.2032	87–27	17	0.0608	63–44
	5	0.1847	81–29	18	0.0532	62–45
	6	0.1691	77–31	19	0.0459	61–48
	7	0.1554	75–32	20	0.0386	61–48
	8	0.1430	74–33	21	0.0314	59–54
	9	0.1317	73–33	22	0.0244	58–54
	10	0.1212	73–35	23	0.0174	58–56
	11	0.1113	70–36	24	0.0104	58–57
	12	0.1020	68–37	25	0.0035	58–57
451	13	0.0932	68–40			

检验统计量的分子变为：

$$\left[\sum_{i=1}^{k} a_i(X^{n-i+1} - X^{(i)})\right]^2 = [(0.3751)(97-23) + \cdots + (0.0035)(58-57)]^2$$

$$= [130.63]^2 = 17{,}064$$

分母为：

$$D = \sum_{i=1}^{n} (X_i - \overline{X})^2 = 17{,}698$$

所以检验统计量变为：

$$T_3 = \frac{17{,}064}{17{,}698} = 0.9642$$

T_3 的值介于分布的 0.10 与 0.50 分位数之间，用表 A17 进行插值得到 p-值，大约为 0.29.

为了找到更精确的 p-值，从表 A18 中获取 $n=50$ 的系数；$b_{50}=-7.677$，$c_{50}=2.212$，且 $d_{50}=0.1436$，并把 T_3 的值带入下式，得：

$$G = b_{50} + c_{50}\ln\left(\frac{T_3 - d_{50}}{1 - T_3}\right)$$

$$= -7.677 + (2.212)\ln\left(\frac{0.9642 - 0.1436}{1 - 0.9642}\right) = -0.7488$$

从表 A1 可知，G 值对应于 $p=0.227$. 这一值比我们刚才用插值得到的 p-值要精确.

■

Shapiro-Wilk 检验所蕴涵的理论太多不能在此一一列举，但是有兴趣的读者可以参看 Shapiro 和 Wilk（1965，1968）的原始文章. 一般说来，如果样本来自正态总体，那么次序统计量与一些常数 a_i 的 Pearson 相关系数 γ 的平方（即检验统计量 T_3）接近 1.0. Stephens（1975）努力拓展已知的表格，但显然没有扩大到我们现在知道的程度. Hartley 和 Pfaffenberger（1972），Bowman 和 Shenton（1975），Pearson，D'Agosino 和 Bowman(1975)提出了一些新的正态性拟合优度检验的方法.

Shapiro-Wilk 检验的一个很有用的特征就是，可以把几个独立的拟合优度检验结果合并成一个来检验其正态性. 这在某些情况下是非常方便的，如一些小样本可能来自不同的总体，但它们本身又不足以拒绝正态性假设，但如果结合在一起却足以拒绝正态性假设.

这种把许多独立的结果结合在一起研究的方法称为 *meta-分析*（meta-analysis），它包括把每个研究中的检验统计量转化为一个标准正态随机变量，或者通过把 p-值 452
转化为正态分布的分位数，或者如 Shapiro-Wilk 检验中使用表 A18 一样，直接通过逆来转化. 然后把这些正态随机变量加起来，除以样本数的平方根，即得到零分布为标准正态分布的检验统计量. Wolf(1986)提倡使用这种方法. 下面的例子将说明如何使用这种方法.

例 6.2.4

当一个近海油气井开始招标时，通常会有许多石油公司参与投标，以争取这个区域的石油开采权. 每个油气井的投标商的数量服从对数指数分布，也就是投标商数量的对数服从正态分布. 然而不同的油气井之间服从的分布的均值与方差各不相同. 同时，任何一口油气井上的投标商通常太少以至不能确定其对数正态假设是否合理.

为了检验假设：

H_0: 投标商的数量服从对数正态分布

对备择假设：它们不是服从对数正态分布. 我们以 16 口不同油气井的投标商数量为样本，对每口油气井的投标商数的对数分别作 Shapiro－Wilk 检验，结果在 $\alpha=0.05$ 的水平下，16 口油气井中有 4 块拒绝零假设. 然而，有些却显然与零假设十分的吻合，其 p-值大于 0.50. 因此，为了综合 16 个检验的结果，我们采取以下的步骤：

1. 把每个 T_3 值按表 A18 转化成为 G 值.

2. 把 16 个 G 值加在一起.

3. 用上面的结果上除以$\sqrt{n}$，得到 Z，它的分布在零假设成立的情形下为近似标准正态.

4. 如果 Z 的值小于表 A1 的 α 分位数，那么我们以水平 α 拒绝零假设.

对于这口油气井的计算结果如下：

453

油气井	投标商数	T_3	G
1	14	0.9243	−0.6550
2	14	0.9757	1.3559
3	14	0.9717	1.0939
4	14	0.8772	−1.5848
5	14	0.9537	0.2345
6	15	0.9135	−1.0093
7	15	0.8629[a]	−1.9321
8	15	0.8786[a]	−1.6806
9	15	0.8515[a]	−2.1011
10	15	0.9226	−0.7966
11	15	0.9581	0.3354
12	15	0.9625	0.5344
13	16	0.9178	−1.0151
14	16	0.8596[a]	−2.1011
15	15	0.9603	0.4323
16	16	0.9669	0.6795
		Total	−8.2099

a在 $\alpha = 0.05$ 水平下显著.

$$Z = \frac{-8.2099}{\sqrt{16}} = -2.0525$$

Z 的值小于从表 A1 得到的 −1.6449，所以在水平 $\alpha = 0.05$ 下拒绝 H_0. 从表 A1 中可以看出 p-值等于 0.020. 因此，对数正态的假设未被证实. ■

有一点必须指出的是，如果观测的数据太大，几乎所有的拟合优度检验都将拒绝零假设. 也就是说，现实中的数据并不总是服从任意已知的分布. 然而，这些已知的分布总是在合理的精度范围内与数据“足够的接近”，因此我们能够假设这些数据服从假设的分布. 拟合优度检验是一种探索数据与假设分布的吻合是否足够接近的方法.

习题

1. 下面的数据是随机抽取的 20 只股票的 12 个月的投资收益：

9.1	5.0	7.3	7.4	5.5
8.6	7.0	4.3	4.7	8.0
4.0	8.5	6.4	6.1	5.8
9.5	5.2	6.7	8.3	9.2

用 Lilliefors 检验来检验其正态性零假设.

2. 15 名新生的入学考试成绩如下：

481	620	642	515	740
562	395	615	596	618
525	584	540	580	598

用 Lilliefors 检验来检验其正态性.

3. 对习题 1 中的假设检验问题用 Shapiro-Wilk 检验.

4. 对习题 2 中的假设检验问题用 Shapiro-Wilk 检验. 454

5. 某商店经理想检验顾客随机到达商店的假设，因此，一天早上，她记录下了连续到达的顾客间的间隔时间如下：

3.6	6.2	12.7
14.2	38.0	3.8
10.8	6.1	10.1
22.1	4.2	4.6
1.4	3.3	8.2

试检验零假设：这些间隔时间服从指数分布.

6. 一段州际高速公路间的特殊路段一个月发生了 20 次事故，下面是事故发生地之间的 19 个距离（英里）：

0.3	6.1	4.3	3.3	1.9
4.8	0.3	1.2	0.8	10.3
1.2	0.1	10.0	1.6	27.6
12.0	14.2	19.7	15.5	

这些事故是随机地发生在这段公路上的吗?

7. 通常情况下，我们假设水流量（流经某河段的水量）数据服从对数指数分布. 为了验证这个假设，收集了不同大小的 8 条河流的水流量数据. 数据包括每星期测量一次的流量（立方英尺/秒），且对每条河流测量星期的多少不一. 用 Shapiro-Wilk 检验来检验测量数据对数的正态性，结果如下：

河流号	测量的周数	T_3的值
1	8	0.972
2	10	0.858
3	6	0.875
4	14	0.840
5	9	0.966
6	10	0.924
7	14	0.881
8	12	0.868

综合以上结果，是否表明流量数据服从对数正态分布?

8. 通常认为年降水量服从正态分布，收集了美国 10 个城市的数据来检验此假设. 对年降水量用 Shapiro-Wilk 检验分析，得到如下的结果：

城市	记录的年数	T_3 的值
1	18	0.875
2	34	0.874
3	26	0.948
4	43	0.980

455

5	40	0.937
6	29	0.915
7	35	0.915
8	38	0.890
9	42	0.963
10	47	0.941

综合以上结果，是否表明年降水量服从正态分布？

6.3 两组独立样本的检验

当两组样本抽自可能不同的总体时，试验者想知道这两个总体分布是否相同．这一节中我们将要描述在这种情形下十分有用的一些检验．当然，其他的一些检验，譬如，中位数检验，Mann-Whitney 检验，或者参数 t 检验也是合适的，它们对于两个总体的中位数或均值差异非常敏感，但它们可能不能检测出其他类别的差异，譬如方差的差异．本节要讲述的两个双边检验的一个优点在于，对两个分布函数间的各种类型的差异，它们是相合的（即能检测出差异）．

首先要讲的检验是 Smirnov 检验（Smirnov, 1939）．它是在 6.1 节中讲述的 Kolmogorov 检验的两样本时的情形，因此有时也称之为 Kolmogorov-Smironov 两样本检验，相应地把 Kolmogorov 检验有时称为 Kolmogorov-Smironov 一样本检验，Smirnov 检验有单边与双边的情形．另一个要讲的双边检验是 Cramér-von Mises 两样本检验．它比 Smirnov 检验稍微难计算一些，但有许多人喜欢用它，因为它似乎能更有效地利用数据．实际上，两种检验在功效上差异很小．

▶Smirnov 检验

数据 数据包含有两组独立的样本，一组容量为 $n, X_1, X_2, \cdots, X_n$；另一组容量为 m，$Y_1, Y_2, \cdots, Y_m$．用 $F(x)$ 与 $G(x)$ 分别代表它们未知的分布函数．

假定条件

1. 样本是随机样本．
456 2. 两组样本是相互独立的．
3. 度量尺度至少是须序的．
4. 为使这个检验成为精确的，假设所用到的随机变量为连续的．

如果随机变量是离散的，检验依然是正确的，但是变得保守（见 Noether, 1967a）．

检验统计量 记 $S_1(x)$ 为基于样本 $X_1, X_2, \cdots, X_n$ 的经验分布函数，$S_2(x)$ 为基于样本 $Y_1, Y_2, \cdots, Y_m$ 的经验分布函数．对应于不同的假设集 A, B, C，相应的检验统计量定义如下：

A.（双边检验）记检验统计量 T_1 定义为两个经验分布函数的最大垂直距离：

$$T_1 = \sup_x |S_1(x) - S_2(x)| \tag{1}$$

B.（单边检验）记检验统计量为 T_1^+ 定义为 $S_1(x)$ 在 $S_2(x)$ 之上的最大垂直距离，也就是：

$$T_1^+ = \sup_x [S_1(x) - S_2(x)] \tag{2}$$

C.（单边检验）记检验统计量为 T_1^- 定义为 $S_2(x)$ 在 $S_1(x)$ 之上的最大垂直距离，也就是：

$$T_1^- = \sup_x [S_2(x) - S_1(x)] \tag{3}$$

零分布 为了求得 T_1, T_1^+, T_1^- 的精确零分布，我们考虑在零假设成立时，X 与 Y 的每种排序是等可能的．正如 Mann-Whitney 检验一样，对每一种排序，计算 T_1, T_1^+, T_1^-．表 A19 与表 A20 分别给出了零分布在 $m=n$ 与 $m \neq n$ 两种情形下的分位数．

在等样本容量的情形下，即 $m=n$ 时，T_1^+ 与 T_1^- 的精确分布为

$$F(x) = 1 - \frac{\binom{2n}{n+c}}{\binom{2n}{n}} \tag{4}$$

其中，c 是小于 $x \cdot n$ 的最大整数．

假设 457

A.（双边检验）

$$H_0: F(x) = G(x) \qquad 对所有 x \in (-\infty, +\infty)$$
$$H_1: F(x) \neq G(x) \qquad 至少对于某个 x$$

如果 T_1 值超过了双边检验情形下的 $1-\alpha$ 分位数，则我们以水平 α 拒绝 H_0，表 A19 和表 A20 分别给出了 $m=n$ 和 $m \neq n$ 的情形下的 $1-\alpha$ 分位数．对于大样本情形，虽然表中未给出，但表的最后采用了近似的方法．近似 p-值可以通过用合适的表进行插值得到．若 $m=n$，更精确的 p-值为 2 倍精确的单边 p-值．

$$单边\ p\text{-}值 = \frac{\binom{2n}{n+nt}}{\binom{2n}{n}} \tag{5}$$

这里的 t 是检验统计量的观测值．

B.（单边检验）

$$H_0: F(x) \leqslant G(x) \qquad 对所有 x \in (-\infty, +\infty)$$
$$H_1: F(x) > G(x) \qquad 至少对于某个 x$$

备择假设有时表述为，“X 倾向于小于 Y”，这比位置备择假设：“X 与 Y 只有位置参数（均值或中位数）不同”的表述更一般．

如果 T_1^+ 值超过了单边检验情形下的 $1-\alpha$ 分位数，则我们以水平 α 拒绝 H_0，表 A19 和表 A20 分别给出了 $m=n$ 和 $m \neq n$ 的情形下的 $1-\alpha$ 分位数．近似 p-值可以通过用合适的表进行插值得到．若 $m=n$，精确的 p-值可以通过（5）式得到．

C.（单边检验）

$$H_0: F(x) \geqslant G(x) \quad 对所有\ x \in (-\infty, +\infty)$$
$$H_1: F(x) < G(x) \quad 至少对于某个\ x$$

如果我们怀疑 X 漂移到了 Y 的右边（也就是 X 比 Y 大），那么我们采用这个单边检验.

如果 T_1^- 值超过了单边检验情形下的 $1-\alpha$ 分位数，则我们以水平 α 拒绝 H_0，表 A19 和表 A20 分别给出了 $m=n$ 和 $m \neq n$ 的情形下的 $1-\alpha$ 分位数. 近似 p-值可以通
458 过用合适的表进行插值得到. 若 $m=n$，精确的 p-值可以通过（5）式得到.

计算机辅助　*StatXact* 可做这种检验，在软件里称它为 Kolmogorov-Smirnov 两样本检验. 并且只要可能，就可以计算精确 p-值. ◀

例 6.3.1

从某个总体中抽出一组容量为 9 的随机样本 $X_1, X_2, \cdots, X_9$，从另一个总体中抽出另一组容量为 15 的随机样本 $Y_1, Y_2, \cdots, Y_{15}$. 它们的经验分布函数如图 6-8

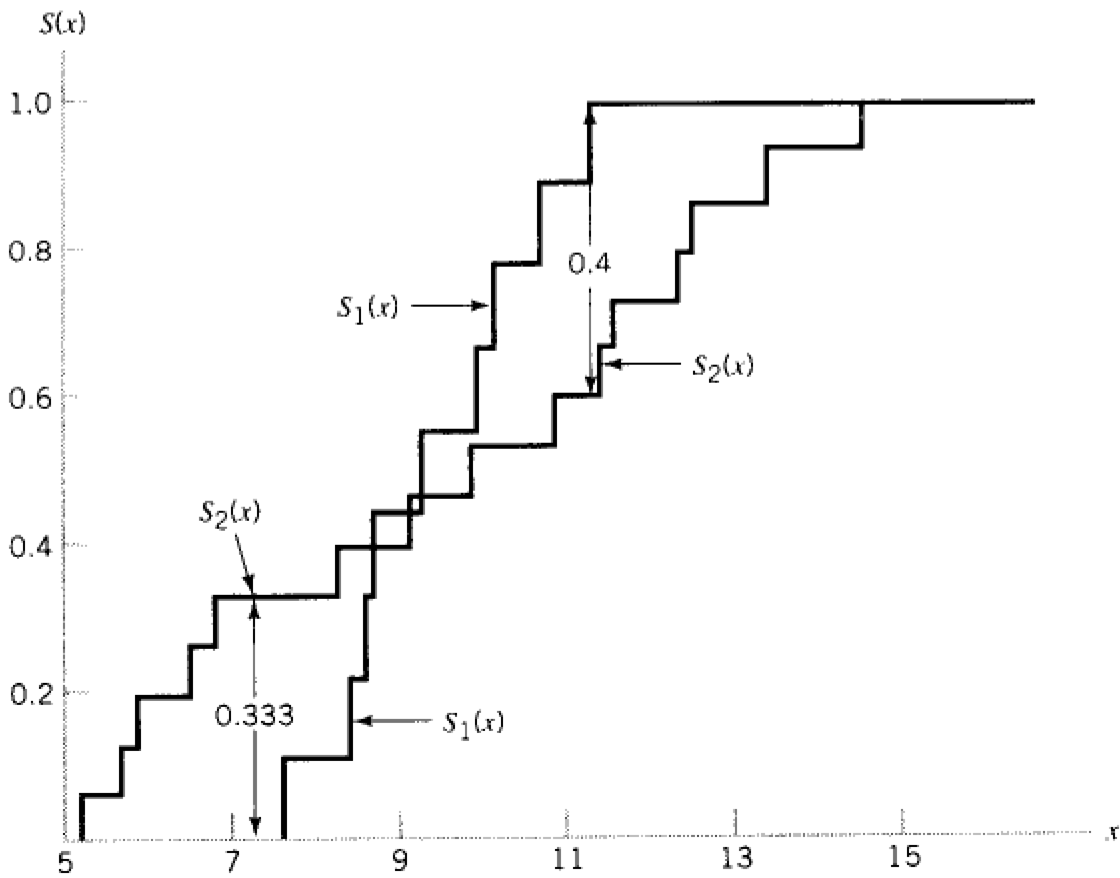
零假设是这两个总体有相同的分布函数. 如果记它们的分布函数分别为 $F(x)$，$G(x)$，那么零假设可写为：

$$H_0: F(x) = G(x) \quad 对所有\ x \in (-\infty, +\infty)$$

相应的备择假设为

$$H_1: F(x) \neq G(x) \quad 至少对于某个\ x$$

为方便起见，我们把样本从小到大排序. 它们的值以及其它们的经验分布函数的有关信息如下给出：



459 图 6-8　X 的经验分布函数 $S_1(x)$，Y 的经验分布函数 $S_2(x)$，以及它们的最大垂直距离

X_i	Y_i	$S_1(x)-S_2(x)$	X_i	Y_i	$S_1(x)-S_2(x)$
	5.2	$0-1/15=-1/15$		9.8	$5/9-8/15=1/45$
	5.7	$0-2/15=-2/15$	9.9		$6/9-8/15=2/15$
	5.9	$0-3/15=-1/5$	10.1		$7/9-8/15=11/45$
	6.5	$0-4/15=-4/15$	10.6		$8/9-8/15=16/45$
	6.8	$0-5/15=-1/3$		10.8	$8/9-9/15=13/45$
7.6		$1/9-5/15=-2/9$	11.2		$1-9/15=2/5$
	8.2	$1/9-6/15=-13/45$		11.3	$1-10/15=1/3$
8.4		$2/9-6/15=-8/45$		11.5	$1-11/15=4/15$
8.6		$3/9-6/15=-1/15$		12.3	$1-12/15=1/5$
8.7		$4/9-6/15=2/4$		12.5	$1-13/15=2/15$
	9.1	$4/9-7/15=-1/15$		13.4	$1-14/15=1/15$
9.3		$5/9-7/15=4/45$		14.6	$1-1=0$

双边检验的检验统计量由（1）式给出，如下，

$$T_1=\sup_x |S_1(x)-S_2(x)|=\frac{2}{5}=0.400$$

它是 $S_1(x)$ 与 $S_2(x)$ 的最大绝对距离，出现在 $x=11.2$ 和 $x=11.3$ 之间. 从图 6-8中可以看出，T_1 的值等于 0.400，而且容易看出，$S_1(x)$ 与 $S_2(x)$ 的距离只有在观测点，$x=X_i$ 或 $x=Y_i$ 上发生变化. 这就是为什么我们只需要计算$S_1(x)-S_2(x)$在观测点处的值的原因.

对于双边检验及 $n=9=N_1$，$m=15=N_2$，我们可以从表 A20 得到 T_1 的 0.95 分位数，为 $w_{0.95}=8/15$. 对于这组数据，$T_1=0.400$，因此在 0.05 的水平上，我们接受 H_0. 从表中我们可以估计出 p-值稍微大于 0.20.

为了便于比较，我们计算出基于渐近分布的 0.95 分位数为

$$w_{0.95}\cong 1.36\sqrt{\frac{m+n}{mn}}\doteq 0.573$$

它比精确值 8/15 =0.533 稍大一些，这表明了利用渐近分布来近似容易产生保守的检验.

注意，这个例子里的许多计算可以省略，因为我们可以通过数据或 $S_1(x)$ 与 $S_2(x)$在图 6-8中的大致图像来观察，在许多 X_i，Y_j 上不可能得到 $|S_1(x)-S_2(x)|$ 的最大值，因此可以省去这部分计算，而只考虑那些比较有可能的 X_i, Y_j 进行计算即可. 460

如果用单边检验适合，可用它代替双边检验，那么从前面的数据表，对于假设 B，检验统计量为

$$T_1^+=\sup_x [S_1(x)-S_2(x)]=\frac{2}{5}=0.400$$

对于假设 C，检验统计量为

$$T_1^-=\sup_x [S_2(x)-S_1(x)]=\frac{1}{3}=0.333$$

从表 A20 可以看出，两种单边检验的 p-值都大于 0.10. ■

□**理论**　尽管初看上去不很显然，但是统计量 T_1，T_1^+，T_1^- 仅取决于 X 与 Y 在联合次序样本中的排序，而不需要知道观测的确切值．举个例子，假设有 3 个 X 样本和 2 个 Y 样本，那么在联合样本中就有 $\binom{5}{2}=10$ 种不同的排序．对于这些排序，分别算出 T_1, T_1^+, T_1^- 的值，如下：

排序	T_1	T_1^+	T_1^-	排序	T_1	T_1^+	T_1^-
$X<X<X<Y<Y$	1	1	0	$X<Y<X<Y<X$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$X<X<Y<X<Y$	$\frac{2}{3}$	$\frac{2}{3}$	0	$Y<X<X<Y<X$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$
$X<Y<X<X<Y$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{6}$	$X<Y<Y<X<X$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
$Y<X<X<X<Y$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$Y<X<Y<X<X$	$\frac{2}{3}$	0	$\frac{2}{3}$
$X<X<Y<Y<X$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$Y<Y<X<X<X$	1	0	1

如果双边检验中的零假设成立，也就是两个分布函数相同，并且在连续随机变量的假设条件下，每一种排序都是等可能的．在 5.1 节中，关于 Mann-Whitney 检验，我们已用同样的观点对此有更详尽的讨论．因此在双边检验中，每一种排序的概率为：

$$\text{概率}=\frac{1}{\binom{m+n}{n}}=\frac{1}{\binom{5}{3}}=\frac{1}{10} \tag{6}$$

而且可以推得如下的概率分布：

$P(T_1=\frac{1}{3})=1/10$　　$P(T_1^+=0)=1/5$　　$P(T_1^-=0)=1/5$

461 $P(T_1=\frac{1}{2})=3/10$　　$P(T_1^+=\frac{1}{6})=1/10$　　$P(T_1^-=\frac{1}{6})=1/10$

$P(T_1=\frac{2}{3})=2/5$　　$P(T_1^+=\frac{1}{3})=1/5$　　$P(T_1^-=\frac{1}{3})=1/5$

$P(T_1=1)=1/5$　　$P(T_1^+=\frac{1}{2})=1/5$　　$P(T_1^-=\frac{1}{2})=1/5$

$P(T_1^+=\frac{2}{3})=1/5$　　$P(T_1^-=\frac{2}{3})=1/5$

$P(T_1^+=1)=1/10$　　$P(T_1^-=1)=1/10$

在 $n=3$ 和 $m=2$ 时，T_1^+ 与 T_1^- 有相同的分布，这并非偶然．事实上，对于任意的 n，m，它们都有相同的分布．为了节约篇幅，当 α 很小时，表 A19 和表 A20 利用了双边检验 T_1 的 $1-\alpha$ 分位数等于单边检验 T_1^+ 的 $1-\alpha/2$ 分位数这一性质．例如，在上面的例子中，$P(T_1\geq 1)$ 等于 2 倍的 $P(T_1^+\geq 1)$，$P(T_1\geq 2/3)$ 等于 2 倍的 $P(T_1^+\geq 2/3)$．但是，$P(T_1\geq 1/2)$ 并不等于 2 倍的 $P(T_1^+\geq 1/2)$．

单边检验中的零分布（即 H_0 为真时，统计量的分布）也是通过上述的方法得到的，因为在单边零假设下，当 $F(x)$ 与 $G(x)$ 相等时，临界域的水平是最大的．如果两组样本有相同的样本容量，就没有必要通过上述方法来获得上侧分位数，因为 Gnedenko 和 Korolyuk（1951）推导出了 T_1, T_1^+, T_1^- 分布，且分布是样本容量 n 的函数．这些分布函数的推导很有趣，而且在读者所学的数学知识范围内，但由于篇幅的原因，我们不在此写出．读者可以参考 Fisz（1963）来得到一个通俗易懂的推导．

对不等样本容量的例子，寻找分位数的方法也是非常基本的．然而，许多人利用路径计数法简化了上述“记流水账”的方法，使得有更多的表格能存在（Harter 和 Owen，1970）．关于 Smirnov 检验的更一般的讨论可以参看 Steck（1969）．当没有精确

的分位数可用时，我们可以利用 Kim（1969）给出的更接近精确分位数的近似. □

Tsao(1954) 建议对 Smirnov 检验进行修改，使它能用到截尾数据上. 所谓截尾数据，就是在 X 与 Y 中，只有小于 $X^{(r)}$ 的 X，Y 能被观测到，这在生存分析试验中时有发生. Tsao(1954) 借助于迭代办法得到的表，把 Smirnov 检验用到了截尾样本上. Conover(1967a) 推导出了 Tsao 统计量分布的解析表达式. Birnbaum 和 Hall (1960)把 Smirnov 检验推广到了 3 组以及更多样本的情形，并对 3 组具有相同样本容量的情形，他们给出了表格. Conover(1965, 1980) 对于 k 组（$k \leqslant 10$）具有相同样本容量的情形给出了表格. Conover(1967b) 介绍了单边多组样本的 Smirnov 检验，同时 Conover(1980)给出了直到 10 组等样本容量情形的表格.

下面要讲的一个检验是 Cramér-von Mises 两样本检验，这种检验只有双边的情形，且计算起来比 Smirnov 检验稍微复杂些. 462

▶Cramér-von Mises 两样本检验

数据 数据包含有两组独立的样本，$X_1, X_2, \cdots, X_n$ 与 $Y_1, Y_2, \cdots, Y_m$. 用 $F(x)$ 与 $G(x)$ 分别代表它们未知的分布函数.

假定条件

1. 样本是随机样本且相互独立.
2. 度量尺度至少是须序的.
3. 随机变量是连续的，如果随机变量是离散的，那么检验可能变得保守.

检验统计量 记 $S_1(x), S_2(x)$ 分别为两组样本的经验分布函数. 检验统计量 T_2 定义为

$$T_2 = \frac{mn}{(m+n)^2} \sum_{\substack{x=X_i \\ x=Y_j}} [S_1(x) - S_2(x)]^2 \tag{7}$$

其中，差平方必须在每个 X_i 与 Y_j 处求和. 显然可以把这个检验统计量写为

$$T_2 = \frac{mn}{(m+n)^2} \left\{ \sum_{i=1}^{n} [S_1(X_i) - S_2(X_i)]^2 + \sum_{j=1}^{m} [S_1(Y_j) - S_2(Y_j)]^2 \right\} \tag{8}$$

零分布 与 Smirnov 检验和 Mann-Whitney 检验一样，为了求检验统计量的精确零分布，我们考虑在零假设成立时，对 X, Y 联合样本的每种排序都是等可能的，且对每一种排序，计算 T_2. 我们将利用渐近分布（$n, m \to \infty$ 时）来近似所有样本容量情形下的分布.

假设

$$H_0: F(x) = G(x) \quad \text{对所有} x \in (-\infty, +\infty)$$
$$H_1: F(x) \neq G(x) \quad \text{至少对于某个 } x$$

如果 T_2 值超过了下表给出的 $1-\alpha$ 分位数 $w_{1-\alpha}$，则我们以水平 α 拒绝 H_0，这些分位数是建立在渐近分布的基础上，对于 m, n 很大的情形是对的，而且即使对于小样本的情形也是相当的精确（见 Burr, 1964）. 463

$$w_{0.10}=0.046 \quad w_{0.50}=0.119 \quad w_{0.90}=0.347$$
$$w_{0.20}=0.062 \quad w_{0.60}=0.147 \quad w_{0.95}=0.461$$
$$w_{0.30}=0.079 \quad w_{0.70}=0.184 \quad w_{0.99}=0.743$$
$$w_{0.40}=0.097 \quad w_{0.80}=0.241 \quad w_{0.999}=1.168$$

这些值取自于 Anderson 和 Darling(1952). 当 $n+m\leqslant 17$ 时, Burr (1964) 给出了精确的分位数. 近似的 p-值可以通过近似分位数插值得到. ◀

例 6.3.2

对于例 6.3.1 的数据, 为了计算检验统计量 T_2, 我们首先计算

$$\sum_{i=1}^{9}[S_1(X_i)-S_2(X_i)]^2=0.459$$

和

$$\sum_{j=1}^{15}[S_1(Y_j)-S_2(Y_j)]^2=0.657$$

那么, 通过 (8) 式, 可算得

$$T_2=\frac{mn}{(m+n)^2}\left\{\sum_{i=1}^{n}[S_1(X_i)-S_2(X_i)]^2+\sum_{j=1}^{m}[S_1(Y_j)-S_2(Y_j)]^2\right\}$$
$$=\frac{(15)(9)}{(24)^2}(0.459+0.657)=0.262$$

则在 $\alpha=0.05$ 的水平上, 我们接受分布相同的零假设, 因为 $T_2=0.262$, 它小于 $w_{0.95}=0.461$ (刚才已给出). p-值大约为 0.18, 这稍微小于 Smirnov 检验对这组数据算出的 p-值. ■

□**理论** Cramér-von Mises 两样本检验统计量的精确分布可以用和 Smirnov 检验统计量一样的方法得到. 在零假设成立的情况下, 不同的排序在 X,Y 联合样本中的出现是等可能的. 统计量 T_2 可以通过有序的联合样本来计算. 对于小样本情形, Anderson(1962) 和 Burr(1963,1964) 利用我们刚才讲述的方法和一些计算技巧, 得到了 T_2 的精确分位数. □

Fisz (1960) 简单地介绍了统计量 T_2, 他认为这个统计量属于 Lehmann (1951),
464 统计量的渐近分布由 Rosenblatt (1952) 得到. 但 Lehmann 和 Rosenblatt 所研究的统计量为

$$T_3=\frac{m}{2(m+n)}\sum_{i=1}^{n}\left[\frac{R(X^{(i)})}{m}-i\frac{m+n}{mn}\right]^2+\frac{n}{2(m+n)}\sum_{j=1}^{m}\left[\frac{R(Y^{(j)})}{n}-j\frac{m+n}{mn}\right]^2 \tag{9}$$

这与 Fisz 的统计量 T_2 不同, 除非 $m=n$. Fisz 证明了 T_2 与 T_3 有相同的渐近分布, 其中, Rosenblatt 证明了 T_3 与 Cramér-von Mises 拟合优度检验统计量有相同的渐近分布. 因此, 实际上, T_2 的渐近分布是由 Anderson 和 Darling(1952) 在 Cramér-von Mises 拟合优度检验统计量的论文所得到的. 这就是为什么把 T_2 称为 Cramér-von Mises 两样本检验统计量的原因, 相信既不是 Cramér 也不是 von Mises 给出了它的发明.

关于圆周上点的观测所设计的两样本检验，可参看 Stephens(1965b), Maag(1966)以及 Maag 和 Stephens(1968). Bickel(1969) 描述了多元 Smirnov 检验. Fine(1966)考虑了 Cramér-von Mises 统计量，同时，Csörgö(1965) 与 Percus 和 Percus(1970)讨论了 Smirnov 检验的变异. 关于 Smirnov 检验渐近效率的论文参见 Capon(1965), Ramachandramurty(1966), Andel(1967)以及 Klotz(1967).

Rothman 和 Woodroofe(1972), Rao, Schuster 和 Littell(1975) 把这些检验用来检验分布的对称性. Gail 和 Green(1976a)为单边检验提供了更多的表格，并在另一篇论文(Gail 和 Green, 1976a)中了讨论了一个检验的很有趣的应用. 更多理论方面的讨论可以参看 Takacs(1971)与 Kalish 和 Mikulski(1971)的论文.

习题

1. 检验零假设 $F(x) \leqslant G(x)$，其中，来自 $F(x)$的观测为 0.6，0.8，0.8，1.2 和 1.4. 而来自 $G(x)$的观测为 1.3，1.3，1.8，2.4 和 2.9.
2. 从一个社区随机抽取 5 名六年级的学生，对他们进行读写能力的测试，最后得分为 82，74，87，86，75. 再从另外一个社区随机抽取 8 名六年级的学生，对他们进行同样的测试，最后的得分为 88，77，91，88，94，93，83，94. 从测试来看，这两组六年级的学生的读写成绩有差别吗（用 Smirnov 检验）?
3. 对于习题 2 中的数据，用 Cramér-von Mises 检验，并把得到的结果与 Smirnov 检验得到的结果进行比较.
4. 从一些自愿参加的高血压患者中随机地抽取一部分参加治疗 A，这是一种采用药物治疗的办法；另一部分参加治疗 B，这是包括低盐的饮食和有规律的运动. 看通过 6 个月后患者血压的变化，问这两种治疗方法的疗效有区别吗? 465

疗法	舒张血压的变化
A	−14, −62, −38, −19, −21, −28, −32, −40
B	−51, −31, +14, −12, −27, −38, −10, +6

5. 20 位家庭业主参加了一个节约能源的学习，将他们一部分随机分配到项目 A，通过日常生活慢慢地灌输节约能源的习惯，另一部分分配到项目 B，就是在他们的楼顶装一层 6 英寸厚的绝热材料. 下面是他们 12 个月节约能源情况的数据：

家庭业主	项目	能源节约（美元）	家庭业主	项目	能源节约（美元）
1	A	\$143	11	B	\$175
2	A	106	12	B	142
3	B	182	13	B	111
4	B	158	14	A	82
5	B	161	15	A	12
6	A	108	16	A	58
7	B	131	17	A	42
8	A	138	18	B	96
9	A	101	19	B	90
10	A	83	20	B	144

问这两个项目的效果有区别吗?

6. 向 15 只实验鼠注入癌细胞，用来研究所提出的治疗方案的效果. 其中 6 只老鼠进行了预防治疗，另外 9 只老鼠注入了安慰剂（用来控制实验）. 在数个月后测量每只老鼠体内的瘤的大小，数据如下

	瘤的大小
治疗	0.8, 0.0, 0.6, 1.1, 1.2, 0.5
控制	0.6, 1.6, 1.7, 1.3, 2.2, 1.5, 0.7, 0.7, 1.6

问这种治疗在最终减小瘤的大小上有效吗?

思考题

1. 用文中对 $n=3, m=2$ 所得出的 T_1 的精确分布，求出 0.80，0.90，0.95 分位数. 并把这些分位数与表 A20 中的分位数作比较，解释它们的差异.
2. 对 $n=3$，$m=3$，求 T_1，T_1^+，T_1^- 和 T_2 的精确分布.

466 3. 对 $n=m=30$ 与 $n=m=10$，试比较精确的 0.95 分位数与基于渐近分布的近似分位数.

6.4　第 1 章至第 6 章复习题

1. 为了检测一种特殊绳子能承受的最大拉力（断裂点），对 10 根绳子进行了测试. 它们所能承受的最大拉力（磅）如下：780，620，910，900，730，700，630，690，730，840.
   (a) 画出它的经验分布函数图.
   (b) 画出总体分布函数的 90% 的置信界.
   (c) 求总体中位数的近似 90% 的置信区间.
2. 某个城市有 5 个行政区，从每个行政区中随机地抽出 10 个房子，并依照房子及庭院被损坏的程度，给出从 0 到 100 的分数.（0 = 未被损坏，100 = 没有重修的社会价值），下面是所得数据的结果：

房子	行政区 1	行政区 2	行政区 3	行政区 4	行政区 5
1	08	74	92	03	37
2	45	42	79	09	28
3	43	77	99	22	42
4	64	09	38	06	44
5	03	32	31	26	01
6	85	66	83	20	32
7	74	16	27	56	65
8	48	45	76	20	02
9	19	15	82	04	80
10	57	24	37	29	93

   对这组数据，列出所有的你能够用来检验行政区与行政区不同的非参数检验的方法，详细地说明每种检验的优缺点. 选出你认为最好的一种检验，对零假设行政区与行政区没有差异进行检验.
3. Gwen 与 Rich 教同一门课程的不同部分. 当课程结束时，比较他们给学生的等级评分，看看他们所给等级分数的分布是否基本相同：

	A	B	C	D和F
Gwen	14	28	17	3
Rich	6	22	23	7

用χ^2 检验来检验差异是否显著.

4. 对于习题3中的数据采用 Kruskal-Wallis 检验，并与习题3的结果比较. 想想在什么情况下你推荐使用χ^2 检验，什么情况下推荐使用 Kruskal-Wallis 检验?

5. (a) 判断下列说法正确与否：
   (1) 如果两个事件互不相容，那么它们互相独立.
   (2) 当样本量变大时，相合性检验的功效趋近于 α.
   (3) 对于小样本，A. R. E 是相对效率的很好的近似. 467
   (4) 当样本来自双指数分布时，符号检验比 Wilcoxon 符号秩检验更有功效.
   (5) H_0 成立的情况下，秩统计量的精确分布总是可以由简单随机方法求得.
   (6) 如果出现很多的结，我们应该使用中位数检验代替 Kruskal-Wallis 检验.

   (b) 用词填空：
   (1) ____________是临界域的水平.
   (2) ____________是样本空间的子集.
   (3) ____________是试验所有可能结果的集合.
   (4) 如果试验产生了____________结果，那么我们拒绝零假设.
   (5) 拒绝零假设的最小的显著性水平，称为____________.
   (6) 拒绝一个错误的零假设的概率称为____________.

6. 掷一枚不均匀的硬币6次，检验 $H_0:P(H)=1/3$ 对 $H_1:P(H)\neq 1/3$. 如果结果是“所有出现的都是反面朝上”或者“正面朝上的次数多于4次”，那么拒绝零假设 H_0.
   (a) H_0 是简单还是复合假设?
   (b) H_1 是简单还是复合假设?
   (c) 列出在临界域中的点.
   (d) α 的值是多少?
   (e) 写出功效函数的表达式.

7. 一位经济学家计算了每月“景气指数”，最近的24个月的值为：123.6, 121.0, 124.1, 123.4, 125.7, 129.0, 126.8, 127.1, 127.3, 126.7, 124.8, 125.9, 124.7, 125.9, 125.6, 126.0, 125.7, 127.3, 127.7, 129.0, 128.2, 127.9, 127.8, 127.1. 这些数字是否表明“景气指数”有某些趋势?

8. 在某一特定区域，每年给一次“年度最高水位”报告，下面是16年的数据（英尺）：7.4, 7.8, 6.9, 8.1, 8.0, 7.1, 7.4, 6.8, 6.9, 7.6, 7.6, 8.0, 8.3, 7.5, 7.8, 7.1. 零假设为：年度最高水位中位数小于8.0英尺，检验该假设，并求年度最高水位中位数的90%置信区间.

9. 掷骰子60次得到如下结果：

显示的点数	1	2	3	4	5	6
出现次数	12	10	14	8	9	7

检验假设：骰子是均匀的，即每面有相同的出现概率.

10. 从刚毕业参加工作的学生中随机地抽取 90 名，其中 30 名来自于人文与艺术学院，30 名来自于农林学院，30 名来自于工学院. 有一半学生的月工资高于 2000 美元，另一半学生的月工资低于 2000 美元. 在那些月工资高于 2000 美元的人中，9 名来自农林学院，17 名来自人文与艺术学院，19 名来自工学院. 检验假设：3 个学院毕业生月工资的中位数相同.

11. 100 人参与了品尝新牌子的止咳糖浆，并说出什么牌子的比现有的普通的止咳糖浆味道要
468 好、而什么牌子却不如此. 如下表所示，有 15 名受试者认为所有这 4 种新牌子的味道要比现有的好；有 3 名受试者认为 A,B,C 这 3 个牌子比现有的要好，而 D 却不然；等等. 检验零假设：对这 4 种新牌子的口味偏好没有显著差异.

牌子				对应的受试者人数
A	B	C	D	
1	1	1	1	15
1	1	1	0	3
1	1	0	1	3
1	0	1	1	6
0	1	1	1	21
1	1	0	0	1
1	0	1	0	1
0	1	1	0	1
1	0	0	1	2
0	1	0	1	2
0	0	1	1	19
1	0	0	0	3
0	1	0	0	3
0	0	1	0	2
0	0	0	1	13
0	0	0	0	5
				100

12. 下面的数据是对被切除垂体的老鼠不注射或注射不同剂量的肾上腺皮层荷尔蒙（4 种方式处理）后，老鼠的生存天数：

A (不注射)	B	C	D
3	2	4	13
2	1	4	4
2	2	3	6
3	6	4	8
5	14	6	19
4	7	5	19
2	15	4	12
2	2	3	1
3	1	4	4
	4	5	4
			1
			12

检验零假设：这 4 种情况下的处理效果没有差异.

13. 下面的数据代表的是 4 种不同种类的麦子在 13 块不同的地块上的产量： 469

	麦子种类			
地块	A	B	C	D
1	43.60	24.05	19.47	19.41
2	40.40	21.76	16.61	23.84
3	18.08	14.19	16.69	16.08
4	19.57	18.61	17.78	18.29
5	45.20	29.33	20.19	30.08
6	25.87	25.60	23.31	27.04
7	55.20	38.77	21.15	39.95
8	55.32	34.19	18.56	25.12
9	19.79	21.65	23.31	22.45
10	46.24	31.52	22.48	29.28
11	14.88	15.68	19.79	22.56
12	7.52	4.69	20.53	22.08
13	41.17	32.59	29.25	43.95

检验假设：不同种类的麦子产量没有差异.

14. 下面的数据是 180 只老鼠从接种 3 种不同类型的伤寒菌微生物到死亡的天数：

	到死亡的天数												
伤寒菌类型	2	3	4	5	6	7	8	9	10	11	12	13	14
9D	10	8	18	16	3	4	1						
11C	1	3	3	6	6	14	11	4	6	2	3	1	
DSC1	1	2	1	3	8	11	10	7	7	3	4	2	1

例如，有 10 只接种了 9D 型的老鼠在第 2 天就死亡了，等等. 那么老鼠对于不同类型的伤寒菌微生物的反应是否有显著差异？

15. 在习题 14 中，我们假设老鼠从被接种 9D 微生物到死亡的天数服从正态分布是否合理？

16. 下面是 12 名男性和 12 名女性能忍受痛苦的极限值：

男性	8.5	7.9	6.7	7.4	7.5	8.6	8.0	8.1	7.2	8.0	7.8	7.8
女性	6.4	7.8	7.1	8.0	6.6	7.3	8.1	7.4	8.3	8.9	7.8	7.7

检验均值是否相等，检验方差是否相等. 470

17. 下面是 20 个有代表性的公司的普通股票在 1951 到 1952 年间的净收入：

1951	1952	1951	1952
\$1.68	\$1.71	\$4.64	\$4.79
1.72	2.17	4.76	4.33
2.50	2.25	5.35	6.05
2.90	2.43	5.81	7.09
3.11	2.32	6.11	6.38
3.35	3.15	6.35	6.00
3.80	3.30	6.69	6.01
3.85	5.52	8.41	7.41
3.89	3.32	8.83	9.33
4.36	3.76	8.97	9.25

问从 1951 年到 1952 年收入是否有统计意义上的显著上涨？

18. 在一个大学二年级的统计班里，10名自愿受试者参加两个考试，一个是考察数学基础知识，另一个是目前所学的课程，所得结果如下：

学生	数学成绩	目前课程成绩
1	37	23
2	44	34
3	55	59
4	70	25
5	26	16
6	39	12
7	26	16
8	30	25
9	85	60
10	83	69

请问，这两个考试的成绩是否有显著的相关性？

19. 在实验室的条件下和给定的时间内，对4种不同类型的轮胎进行测试，每种类型的轮胎选10个. 试验结束后，测量了轮胎的平均磨损深度（cm），得到结果如下：

轮胎	类型 1	类型 2	类型 3	类型 4
1	0.34	0.18	0.40	0.33
2	0.31	0.31	0.21	0.29
3	0.08	0.16	0.27	0.13
4	0.26	0.00	0.38	0.24
5	0.29	0.07	0.00	0.10
6	0.00	0.12	0.08	0.45
7	0.09	0.00	0.19	0.37
8	0.14	0.00	0.36	0.19
9	0.26	0.04	0.34	0.53
10	0.19	0.09	0.44	0.56

471

我们的主要兴趣是，找出哪种（或者哪几种）轮胎的磨损较小，如果有，是哪种（或者哪几种）？

20. 记 X 为一个家庭中拥有的汽车的数量，记 Y 为一个家庭中拥有驾照的司机的数量. 下表给出了人群中出现各种 X 与 Y 值的频率.

X	Y	(X, Y)的概率	X	Y	(X, Y)的概率
0	0	0.10	1	2	0.10
0	1	0.10	2	0	0.10
0	2	0.05	2	1	0.10
1	0	0.20	2	2	0.05
1	1	0.20			

(a) 求 $E(X)$ 与 $E(Y)$.

(b) 求 X 的中位数.

(c) 求 Y 的方差.

(d) 问 X 与 Y 是否独立？

(e) 画出 Y 的分布函数图.

(f) 如果 $f(x,y)$ 是 (X,Y) 的概率函数，求 $f(1,1)$.

(g) 如果 $F(x,y)$ 是 (X,Y) 的分布函数，求 $F(1,1)$.

(h) 求 X 与 Y 的相关系数.

21. 下面的数据是家庭拥有汽车与有驾照的司机的随机样本，其中 X 与 Y 的含义同习题 20，也就是，有 3 个家庭既没有汽车也没有拥有驾照的司机，等等.

X	Y	频数	X	Y	频数
0	0	3	1	2	1
0	1	2	2	0	2
0	2	0	2	1	0
1	0	3	2	2	2
1	1	4			

(a) 求 X 与 Y 的样本均值.

(b) 求 X 的样本中位数.

(c) 求 Y 的样本方差(用 n 作分母,而不是 $n-1$).

(d) 画出 Y 的经验分布函数图.

22. 7 个成年人参加了一个旨在提高阅读速度的培训班，在参加培训班前的阅读速度（字/分钟）与参加培训班后的阅读速度，如下表所示：

之前	270	250	185	310	200	260	260
之后	390	380	310	470	380	400	510

472

求通过这个阅读班人们期望得到阅读速度平均增加的 95% 的置信区间.

23. 比较两只黑猩猩，Dick 和 Jane，看谁能更好地操作并找到"正确"键. 令 X 代表 Dick 在按到"正确"键前按键的次数，总共按了 30 次；Y 代表 Jane 在按到"正确"键前按键的次数，总共按了 20 次. 从下面的表中是否可以说明某只黑猩猩比另外一只更好地找到"正确"键.

Dick		Jane	
X	频数	Y	频数
1	12	1	12
2	8	2	7
3	7	3	1
4	3	4	0
共计	30	共计	20

24. 一个摩托车手想看看用无铅优化汽油是否比普通的无铅汽油能跑更多的里程数. 他每次加油前抛一次硬币，如果正面朝上，那么就加无铅优化汽油，如果反面朝上，则加普通的无铅汽油. 他每次把汽油箱里的汽油跑完，然后算出一箱汽油所跑的里程. 结果有 3 次加的是普通的无铅汽油，所跑的里程为 21.3，21.2 和 21.6；有 5 次加的是无铅优化汽油，所跑的里程为 22.1，22.7，22.3，21.5，21.8.

(a) 用无铅优化汽油，他的车能跑更长的里程数吗？使用秩检验.

(b) 使用 Fisher 随机化检验来分析数据，并求得精确的 p-值.

(c) 使用 Smirnov 检验来分析数据，并求得精确的 p-值.

25. 15 位妇女的年龄与血压记录如下表：

年龄	血压	年龄	血压
48	144	54	151
60	168	56	152
35	135	31	141
38	125	24	144
55	159	77	170
51	148	63	157
49	128	67	162
38	134		

问年龄和血压有明显的单调关系吗?

26. 用习题 25 中的数据作为随机样本，预测 50 岁妇女的平均血压.

27. 12 名学生参加了一个考试，得到了如下的分数：

62, 74, 82, 84, 86, 86, 89, 90, 94, 94, 95, 97

473 用拟合优度检验来检验这些数据是否来自于正态总体.

附　　表

表 A1　正态分布
表 A2　χ^2 分布
表 A3　二项分布
表 A4　二项参数 p 的精确置信区间
表 A5　当 $r+m=1$ 时，非参数容忍限的样本容量
表 A6　当 $r+m=2$ 时，非参数容忍限的样本容量
表 A7　Mann-Whitney 检验统计量的分位数
表 A8　小样本 Kruskal-Wallis 检验统计量的分位数
表 A9　平方秩检验统计量的分位数
表 A10　Spearman ρ 的分位数
表 A11　Kendall 检验统计量 T 和 Kendall τ 的分位数
表 A12　Wilcoxon 符号秩检验统计量的分位数
表 A13　Kolmogorov 检验统计量的分位数
表 A14　Lilliefors 正态性检验统计量的分位数
表 A15　Lilliefors 指数分布检验统计量的分位数
表 A16　Shapiro-Wilk 检验的系数
表 A17　Shapiro-Wilk 检验统计量的分位数
表 A18　变换 Shapiro-Wilk 统计量到近似正态分布的方法
表 A19　等样本容量 n 的两样本 Smirnov 检验统计量的分位数
表 A20　不等样本容量 n,m 的两样本 Smirnov 检验统计量的分位数
表 A21　t 分布
表 A22　自由度为 k_1,k_2 的 F 分布

表 A1 正态分布[a]

		选择值	$z_{0.0001}=-3.7190$ $z_{0.9999}=3.7190$		$z_{0.0005}=-3.2905$ $z_{0.9995}=3.2905$		$z_{0.025}=-1.9600$ $z_{0.975}=1.9600$		$z_{0.05}=-1.6449$ $z_{0.95}=1.6449$	
p	**0.000**	**0.001**	**0.002**	**0.003**	**0.004**	**0.005**	**0.006**	**0.007**	**0.008**	**0.009**
0.00		−3.0902	−2.8782	−2.7478	−2.6521	−2.5758	−2.5121	−2.4573	−2.4089	−2.3656
0.01	−2.3263	−2.2904	−2.2571	−2.2262	−2.1973	−2.1701	−2.1444	−2.1201	−2.0969	−2.0749
0.02	−2.0537	−2.0335	−2.0141	−1.9954	−1.9774	−1.9600	−1.9431	−1.9268	−1.9110	−1.8957
0.03	−1.8808	−1.8663	−1.8522	−1.8384	−1.8250	−1.8119	−1.7991	−1.7866	−1.7744	−1.7624
0.04	−1.7507	−1.7392	−1.7279	−1.7169	−1.7060	−1.6954	−1.6849	−1.6747	−1.6646	−1.6546
0.05	−1.6449	−1.6352	−1.6258	−1.6164	−1.6072	−1.5982	−1.5893	−1.5805	−1.5718	−1.5632
0.06	−1.5548	−1.5464	−1.5382	−1.5301	−1.5220	−1.5141	−1.5063	−1.4985	−1.4909	−1.4833
0.07	−1.4758	−1.4684	−1.4611	−1.4538	−1.4466	−1.4395	−1.4325	−1.4255	−1.4187	−1.4118
0.08	−1.4051	−1.3984	−1.3917	−1.3852	−1.3787	−1.3722	−1.3658	−1.3595	−1.3532	−1.3469
0.09	−1.3408	−1.3346	−1.3285	−1.3225	−1.3165	−1.3106	−1.3047	−1.2988	−1.2930	−1.2873
0.10	−1.2816	−1.2759	−1.2702	−1.2646	−1.2591	−1.2536	−1.2481	−1.2426	−1.2372	−1.2319
0.11	−1.2265	−1.2212	−1.2160	−1.2107	−1.2055	−1.2004	−1.1952	−1.1901	−1.1850	−1.1800
0.12	−1.1750	−1.1700	−1.1650	−1.1601	−1.1552	−1.1503	−1.1455	−1.1407	−1.1359	−1.1311
0.13	−1.1264	−1.1217	−1.1170	−1.1123	−1.1077	−1.1031	−1.0985	−1.0939	−1.0893	−1.0848
0.14	−1.0803	−1.0758	−1.0714	−1.0669	−1.0625	−1.0581	−1.0537	−1.0494	−1.0450	−1.0407
0.15	−1.0364	−1.0322	−1.0279	−1.0237	−1.0194	−1.0152	−1.0110	−1.0069	−1.0027	−0.9986
0.16	−0.9945	−0.9904	−0.9863	−0.9822	−0.9782	−0.9741	−0.9701	−0.9661	−0.9621	−0.9581
0.17	−0.9542	−0.9502	−0.9463	−0.9424	−0.9385	−0.9346	−0.9307	−0.9269	−0.9230	−0.9192
0.18	−0.9154	−0.9116	−0.9078	−0.9040	−0.9002	−0.8965	−0.8927	−0.8890	−0.8853	−0.8816
0.19	−0.8779	−0.8742	−0.8705	−0.8669	−0.8633	−0.8596	−0.8560	−0.8524	−0.8488	−0.8452
0.20	−0.8416	−0.8381	−0.8345	−0.8310	−0.8274	−0.8239	−0.8204	−0.8169	−0.8134	−0.8099
0.21	−0.8064	−0.8030	−0.7995	−0.7961	−0.7926	−0.7892	−0.7858	−0.7824	−0.7790	−0.7756
0.22	−0.7722	−0.7688	−0.7655	−0.7621	−0.7588	−0.7554	−0.7521	−0.7488	−0.7454	−0.7421
0.23	−0.7388	−0.7356	−0.7323	−0.7290	−0.7257	−0.7225	−0.7192	−0.7160	−0.7128	−0.7095
0.24	−0.7063	−0.7031	−0.6999	−0.6967	−0.6935	−0.6903	−0.6871	−0.6840	−0.6808	−0.6776
0.25	−0.6745	−0.6713	−0.6682	−0.6651	−0.6620	−0.6588	−0.6557	−0.6526	−0.6495	−0.6464
0.26	−0.6433	−0.6403	−0.6372	−0.6341	−0.6311	−0.6280	−0.6250	−0.6219	−0.6189	−0.6158
0.27	−0.6128	−0.6098	−0.6068	−0.6038	−0.6008	−0.5978	−0.5948	−0.5918	−0.5888	−0.5858
0.28	−0.5828	−0.5799	−0.5769	−0.5740	−0.5710	−0.5681	−0.5651	−0.5622	−0.5592	−0.5563
0.29	−0.5534	−0.5505	−0.5476	−0.5446	−0.5417	−0.5388	−0.5359	−0.5330	−0.5302	−0.5273
0.30	−0.5244	−0.5215	−0.5187	−0.5158	−0.5129	−0.5101	−0.5072	−0.5044	−0.5015	−0.4987
0.31	−0.4959	−0.4930	−0.4902	−0.4874	−0.4845	−0.4817	−0.4789	−0.4761	−0.4733	−0.4705
0.32	−0.4677	−0.4649	−0.4621	−0.4593	−0.4565	−0.4538	−0.4510	−0.4482	−0.4454	−0.4427
0.33	−0.4399	−0.4372	−0.4344	−0.4316	−0.4289	−0.4261	−0.4234	−0.4207	−0.4179	−0.4152
0.34	−0.4125	−0.4097	−0.4070	−0.4043	−0.4016	−0.3989	−0.3961	−0.3934	−0.3907	−0.3880
0.35	−0.3853	−0.3826	−0.3799	−0.3772	−0.3745	−0.3719	−0.3692	−0.3665	−0.3638	−0.3611
0.36	−0.3585	−0.3558	−0.3531	−0.3505	−0.3478	−0.3451	−0.3425	−0.3398	−0.3372	−0.3345
0.37	−0.3319	−0.3292	−0.3266	−0.3239	−0.3213	−0.3186	−0.3160	−0.3134	−0.3107	−0.3081
0.38	−0.3055	−0.3029	−0.3002	−0.2976	−0.2950	−0.2924	−0.2898	−0.2871	−0.2845	−0.2819
0.39	−0.2793	−0.2767	−0.2741	−0.2715	−0.2689	−0.2663	−0.2637	−0.2611	−0.2585	−0.2559
0.40	−0.2533	−0.2508	−0.2482	−0.2456	−0.2430	−0.2404	−0.2378	−0.2353	−0.2327	−0.2301
0.41	−0.2275	−0.2250	−0.2224	−0.2198	−0.2173	−0.2147	−0.2121	−0.2096	−0.2070	−0.2045
0.42	−0.2019	−0.1993	−0.1968	−0.1942	−0.1917	−0.1891	−0.1866	−0.1840	−0.1815	−0.1789
0.43	−0.1764	−0.1738	−0.1713	−0.1687	−0.1662	−0.1637	−0.1611	−0.1586	−0.1560	−0.1535
0.44	−0.1510	−0.1484	−0.1459	−0.1434	−0.1408	−0.1383	−0.1358	−0.1332	−0.1307	−0.1282
0.45	−0.1257	−0.1231	−0.1206	−0.1181	−0.1156	−0.1130	−0.1105	−0.1080	−0.1055	−0.1030
0.46	−0.1004	−0.0979	−0.0954	−0.0929	−0.0904	−0.0878	−0.0853	−0.0828	−0.0803	−0.0778
0.47	−0.0753	−0.0728	−0.0702	−0.0677	−0.0652	−0.0627	−0.0602	−0.0577	−0.0552	−0.0527
0.48	−0.0502	−0.0476	−0.0451	−0.0426	−0.0401	−0.0376	−0.0351	−0.0326	−0.0301	−0.0276
0.49	−0.0251	−0.0226	−0.0201	−0.0175	−0.0150	−0.0125	−0.0100	−0.0075	−0.0050	−0.0025
0.50	0.0000	0.0025	0.0050	0.0075	0.0100	0.0125	0.0150	0.0175	0.0201	0.0226
0.51	0.0251	0.0276	0.0301	0.0326	0.0351	0.0376	0.0401	0.0426	0.0451	0.0476
0.52	0.0502	0.0527	0.0552	0.0577	0.0602	0.0627	0.0652	0.0677	0.0702	0.0728
0.53	0.0753	0.0778	0.0803	0.0828	0.0853	0.0878	0.0904	0.0929	0.0954	0.0979
0.54	0.1004	0.1030	0.1055	0.1080	0.1105	0.1130	0.1156	0.1181	0.1206	0.1231

（续）

p	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.55	0.1257	0.1282	0.1307	0.1332	0.1358	0.1383	0.1408	0.1434	0.1459	0.1484
0.56	0.1510	0.1535	0.1560	0.1586	0.1611	0.1637	0.1662	0.1687	0.1713	0.1738
0.57	0.1764	0.1789	0.1815	0.1840	0.1866	0.1891	0.1917	0.1942	0.1968	0.1993
0.58	0.2019	0.2045	0.2070	0.2096	0.2121	0.2147	0.2173	0.2198	0.2224	0.2250
0.59	0.2275	0.2301	0.2327	0.2353	0.2378	0.2404	0.2430	0.2456	0.2482	0.2508
0.60	0.2533	0.2559	0.2585	0.2611	0.2637	0.2663	0.2689	0.2715	0.2741	0.2767
0.61	0.2793	0.2819	0.2845	0.2871	0.2898	0.2924	0.2950	0.2976	0.3002	0.3029
0.62	0.3055	0.3081	0.3107	0.3134	0.3160	0.3186	0.3213	0.3239	0.3266	0.3292
0.63	0.3319	0.3345	0.3372	0.3398	0.3425	0.3451	0.3478	0.3505	0.3531	0.3558
0.64	0.3585	0.3611	0.3638	0.3665	0.3692	0.3719	0.3745	0.3772	0.3799	0.3826
0.65	0.3853	0.3880	0.3907	0.3934	0.3961	0.3989	0.4016	0.4043	0.4070	0.4097
0.66	0.4125	0.4152	0.4179	0.4207	0.4234	0.4261	0.4289	0.4316	0.4344	0.4372
0.67	0.4399	0.4427	0.4454	0.4482	0.4510	0.4538	0.4565	0.4593	0.4621	0.4649
0.68	0.4677	0.4705	0.4733	0.4761	0.4789	0.4817	0.4845	0.4874	0.4902	0.4930
0.69	0.4959	0.4987	0.5015	0.5044	0.5072	0.5101	0.5129	0.5158	0.5187	0.5215
0.70	0.5244	0.5273	0.5302	0.5330	0.5359	0.5388	0.5417	0.5446	0.5476	0.5505
0.71	0.5534	0.5563	0.5592	0.5622	0.5651	0.5681	0.5710	0.5740	0.5769	0.5799
0.72	0.5828	0.5858	0.5888	0.5918	0.5948	0.5978	0.6008	0.6038	0.6068	0.6098
0.73	0.6128	0.6158	0.6189	0.6219	0.6250	0.6280	0.6311	0.6341	0.6372	0.6403
0.74	0.6433	0.6464	0.6495	0.6526	0.6557	0.6588	0.6620	0.6651	0.6682	0.6713
0.75	0.6745	0.6776	0.6808	0.6840	0.6871	0.6903	0.6935	0.6967	0.6999	0.7031
0.76	0.7063	0.7095	0.7128	0.7160	0.7192	0.7225	0.7257	0.7290	0.7323	0.7356
0.77	0.7388	0.7421	0.7454	0.7488	0.7521	0.7554	0.7588	0.7621	0.7655	0.7688
0.78	0.7722	0.7756	0.7790	0.7824	0.7858	0.7892	0.7926	0.7961	0.7995	0.8030
0.79	0.8064	0.8099	0.8134	0.8169	0.8204	0.8239	0.8274	0.8310	0.8345	0.8381
0.80	0.8416	0.8452	0.8488	0.8524	0.8560	0.8596	0.8633	0.8669	0.8705	0.8742
0.81	0.8779	0.8816	0.8853	0.8890	0.8927	0.8965	0.9002	0.9040	0.9078	0.9116
0.82	0.9154	0.9192	0.9230	0.9269	0.9307	0.9346	0.9385	0.9424	0.9463	0.9502
0.83	0.9542	0.9581	0.9621	0.9661	0.9701	0.9741	0.9782	0.9822	0.9863	0.9904
0.84	0.9945	0.9986	1.0027	1.0069	1.0110	1.0152	1.0194	1.0237	1.0279	1.0322
0.85	1.0364	1.0407	1.0450	1.0494	1.0537	1.0581	1.0625	1.0669	1.0714	1.0758
0.86	1.0803	1.0848	1.0893	1.0939	1.0985	1.1031	1.1077	1.1123	1.1170	1.1217
0.87	1.1264	1.1311	1.1359	1.1407	1.1455	1.1503	1.1552	1.1601	1.1650	1.1700
0.88	1.1750	1.1800	1.1850	1.1901	1.1952	1.2004	1.2055	1.2107	1.2160	1.2212
0.89	1.2265	1.2319	1.2372	1.2426	1.2481	1.2536	1.2591	1.2646	1.2702	1.2759
0.90	1.2816	1.2873	1.2930	1.2988	1.3047	1.3106	1.3165	1.3225	1.3285	1.3346
0.91	1.3408	1.3469	1.3532	1.3595	1.3658	1.3722	1.3787	1.3852	1.3917	1.3984
0.92	1.4051	1.4118	1.4187	1.4255	1.4325	1.4395	1.4466	1.4538	1.4611	1.4684
0.93	1.4758	1.4833	1.4909	1.4985	1.5063	1.5141	1.5220	1.5301	1.5382	1.5464
0.94	1.5548	1.5632	1.5718	1.5805	1.5893	1.5982	1.6072	1.6164	1.6258	1.6352
0.95	1.6449	1.6546	1.6646	1.6747	1.6849	1.6954	1.7060	1.7169	1.7279	1.7392
0.96	1.7507	1.7624	1.7744	1.7866	1.7991	1.8119	1.8250	1.8384	1.8522	1.8663
0.97	1.8808	1.8957	1.9110	1.9268	1.9431	1.9600	1.9774	1.9954	2.0141	2.0335
0.98	2.0537	2.0749	2.0969	2.1201	2.1444	2.1701	2.1973	2.2262	2.2571	2.2904
0.99	2.3263	2.3656	2.4089	2.4573	2.5121	2.5758	2.6521	2.7478	2.8782	3.0902

来源：由 R. L. Iman 产生，并得到应用许可.

a 表中的数是标准正态随机变量 Z 的分位数 z_p，满足 $P(Z\leqslant z_p)=p$ 和 $P(Z>z_p)=1-p$，注意，其中 p 值的前 2 位小数点确定所用的行，第 3 位小数点确定来查到 z_p 的列.

表 A2 χ^2 分布[a]

	$p = 0.750$	0.900	0.950	0.975	0.990	0.995	0.999
$k = 1$	1.323	2.706	3.841	5.024	6.635	7.879	10.83
2	2.773	4.605	5.991	7.378	9.210	10.60	13.82
3	4.108	6.251	7.815	9.348	11.34	12.84	16.27
4	5.385	7.779	9.488	11.14	13.28	14.86	18.47
5	6.626	9.236	11.07	12.83	15.09	16.75	20.51
6	7.841	10.64	12.59	14.45	16.81	18.55	22.46
7	9.037	12.02	14.07	16.01	18.48	20.28	24.32
8	10.22	13.36	15.51	17.53	20.09	21.96	26.13
9	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	13.70	17.28	19.68	21.92	24.73	26.76	31.26
12	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	24.93	29.62	32.67	35.48	38.93	41.40	46.80
22	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	28.24	33.20	36.42	39.37	42.98	45.56	51.18
25	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	31.53	36.74	40.11	43.19	46.96	49.64	55.48
28	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	45.62	51.81	55.76	59.34	63.69	66.77	73.40
50	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	77.58	85.53	90.53	95.02	100.4	104.2	112.3
80	88.13	96.58	101.9	106.6	112.3	116.3	124.8
90	98.65	107.6	113.1	118.1	124.1	128.3	137.2
100	109.1	118.5	124.3	129.6	135.8	140.2	149.4
z_p	0.675	1.282	1.645	1.960	2.326	2.576	3.090

$k>100$，使用近似值 $w_p = \left(\frac{1}{2}\right)(z_p + \sqrt{2k-1})^2$，或者更精确的 $w_p = k\left(1-\frac{2}{9k}+z_p\sqrt{\frac{2}{9k}}\right)^3$，这里 z_p 是标准正态分布的 p 分位数，它列在了表的最下面一行.

来源：从 Pearson 和 Hartley（1976）第一卷表 8 节略，经 *Biometrika* 委托许可使用.

a 表中的数是服从自由度为 k 的 χ^2 分布的随机变量 W 的 p 分位数 w_p，对选择的 p，满足 $P(W\leqslant w_p) = p$ 和 $P(W>w_p) = 1-p$.

表A3 二项分布[a]

n	y	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
1	0	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000	0.4500	0.4000	0.3500	0.3000	0.2500	0.2000	0.1500	0.1000	0.0500
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500	0.2025	0.1600	0.1225	0.0900	0.0625	0.0400	0.0225	0.0100	0.0025
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500	0.6975	0.6400	0.5775	0.5100	0.4375	0.3600	0.2775	0.1900	0.0975
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250	0.0911	0.0640	0.0429	0.0270	0.0156	0.0080	0.0034	0.0010	0.0001
	1	0.9928	0.9720	0.9392	0.8960	0.8438	0.7840	0.7182	0.6480	0.5748	0.5000	0.4252	0.3520	0.2818	0.2160	0.1562	0.1040	0.0608	0.0280	0.0072
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750	0.8336	0.7840	0.7254	0.6570	0.5781	0.4880	0.3859	0.2710	0.1426
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625	0.0410	0.0256	0.0150	0.0081	0.0039	0.0016	0.0005	0.0001	0.0000
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125	0.2415	0.1792	0.1265	0.0837	0.0508	0.0272	0.0120	0.0037	0.0005
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875	0.6090	0.5248	0.4370	0.3483	0.2617	0.1808	0.1095	0.0523	0.0140
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375	0.9085	0.8704	0.8215	0.7599	0.6836	0.5904	0.4780	0.3439	0.1855
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0312	0.0185	0.0102	0.0053	0.0024	0.0010	0.0003	0.0001	0.0000	0.0000
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875	0.1312	0.0870	0.0540	0.0308	0.0156	0.0067	0.0022	0.0005	0.0000
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000	0.4069	0.3174	0.2352	0.1631	0.1035	0.0579	0.0266	0.0086	0.0012
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125	0.7438	0.6630	0.5716	0.4718	0.3672	0.2627	0.1648	0.0815	0.0226
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688	0.9497	0.9222	0.8840	0.8319	0.7627	0.6723	0.5563	0.4095	0.2262
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156	0.0083	0.0041	0.0018	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094	0.0692	0.0410	0.0223	0.0109	0.0046	0.0016	0.0004	0.0001	0.0000
	2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438	0.2553	0.1792	0.1174	0.0705	0.0376	0.0170	0.0059	0.0013	0.0001
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6562	0.5585	0.4557	0.3529	0.2557	0.1694	0.0989	0.0473	0.0158	0.0022
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906	0.8364	0.7667	0.6809	0.5789	0.4661	0.3446	0.2235	0.1143	0.0328
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844	0.9723	0.9533	0.9246	0.8824	0.8220	0.7379	0.6229	0.4686	0.2649
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078	0.0037	0.0016	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625	0.0357	0.0188	0.0090	0.0038	0.0013	0.0004	0.0001	0.0000	0.0000
	2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266	0.1529	0.0963	0.0556	0.0288	0.0129	0.0047	0.0012	0.0002	0.0000
	3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000	0.3917	0.2898	0.1998	0.1260	0.0706	0.0333	0.0121	0.0027	0.0002
	4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734	0.6836	0.5801	0.4677	0.3529	0.2436	0.1480	0.0738	0.0257	0.0038
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375	0.8976	0.8414	0.7662	0.6706	0.5551	0.4233	0.2834	0.1497	0.0444
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922	0.9848	0.9720	0.9510	0.9176	0.8665	0.7903	0.6794	0.5217	0.3017
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

（续）

n	y	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039	0.0017	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352	0.0181	0.0085	0.0036	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000
	2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445	0.0885	0.0498	0.0253	0.0113	0.0042	0.0012	0.0002	0.0000	0.0000
	3	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633	0.2604	0.1737	0.1061	0.0580	0.0273	0.0104	0.0029	0.0004	0.0000
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367	0.5230	0.4059	0.2936	0.1941	0.1138	0.0563	0.0214	0.0050	0.0004
	5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555	0.7799	0.6846	0.5722	0.4482	0.3215	0.2031	0.1052	0.0381	0.0058
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648	0.9368	0.8936	0.8309	0.7447	0.6329	0.4967	0.3428	0.1869	0.0572
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961	0.9916	0.9832	0.9681	0.9424	0.8999	0.8322	0.7275	0.5695	0.3366
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020	0.0008	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195	0.0091	0.0038	0.0014	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898	0.0498	0.0250	0.0112	0.0043	0.0013	0.0003	0.0000	0.0000	0.0000
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539	0.1658	0.0994	0.0536	0.0253	0.0100	0.0031	0.0006	0.0001	0.0000
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000	0.3786	0.2666	0.1717	0.0988	0.0489	0.0196	0.0056	0.0009	0.0000
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461	0.6386	0.5174	0.3911	0.2703	0.1657	0.0856	0.0339	0.0083	0.0006
	6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102	0.8505	0.7682	0.6627	0.5372	0.3993	0.2618	0.1409	0.0530	0.0084
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805	0.9615	0.9295	0.8789	0.8040	0.6997	0.5638	0.4005	0.2252	0.0712
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980	0.9954	0.9899	0.9793	0.9596	0.9249	0.8658	0.7684	0.6126	0.3698
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107	0.0045	0.0017	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547	0.0274	0.0123	0.0048	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719	0.1020	0.0548	0.0260	0.0106	0.0035	0.0009	0.0001	0.0000	0.0000
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770	0.2616	0.1662	0.0949	0.0473	0.0197	0.0064	0.0014	0.0001	0.0000
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230	0.4956	0.3669	0.2485	0.1503	0.0781	0.0328	0.0099	0.0016	0.0001
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281	0.7340	0.6177	0.4862	0.3504	0.2241	0.1209	0.0500	0.0128	0.0010
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453	0.9004	0.8327	0.7384	0.6172	0.4744	0.3222	0.1798	0.0702	0.0115
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893	0.9767	0.9536	0.9140	0.8507	0.7560	0.6242	0.4557	0.2639	0.0861
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9975	0.9940	0.9865	0.9718	0.9437	0.8926	0.8031	0.6513	0.4013
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	0	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.8981	0.6974	0.4922	0.3221	0.1971	0.1130	0.0606	0.0302	0.0139	0.0059	0.0022	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9848	0.9104	0.7788	0.6174	0.4552	0.3127	0.2001	0.1189	0.0652	0.0327	0.0148	0.0059	0.0020	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000
	3	0.9984	0.9815	0.9306	0.8389	0.7133	0.5696	0.4256	0.2963	0.1911	0.1133	0.0610	0.0293	0.0122	0.0043	0.0012	0.0002	0.0000	0.0000	0.0000
	4	0.9999	0.9972	0.9841	0.9496	0.8854	0.7897	0.6683	0.5328	0.3971	0.2744	0.1738	0.0994	0.0501	0.0216	0.0076	0.0020	0.0003	0.0000	0.0000
	5	1.0000	0.9997	0.9973	0.9883	0.9657	0.9218	0.8513	0.7535	0.6331	0.5000	0.3669	0.2465	0.1487	0.0782	0.0343	0.0117	0.0027	0.0003	0.0000
	6	1.0000	1.0000	0.9997	0.9980	0.9924	0.9784	0.9499	0.9006	0.8262	0.7256	0.6029	0.4672	0.3317	0.2103	0.1146	0.0504	0.0159	0.0028	0.0001
	7	1.0000	1.0000	1.0000	0.9998	0.9988	0.9957	0.9878	0.9707	0.9390	0.8867	0.8089	0.7037	0.5744	0.4304	0.2867	0.1611	0.0694	0.0185	0.0016
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9980	0.9941	0.9852	0.9673	0.9348	0.8811	0.7999	0.6873	0.5448	0.3826	0.2212	0.0896	0.0152
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9978	0.9941	0.9861	0.9698	0.9394	0.8870	0.8029	0.6779	0.5078	0.3026	0.1019
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995	0.9986	0.9964	0.9912	0.9802	0.9578	0.9141	0.8327	0.6862	0.4312
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

（续）

n	y	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193	0.0079	0.0028	0.0008	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730	0.0356	0.0153	0.0056	0.0017	0.0004	0.0001	0.0000	0.0000	0.0000
	4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938	0.1117	0.0573	0.0255	0.0095	0.0028	0.0006	0.0001	0.0000	0.0000
	5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872	0.2607	0.1582	0.0846	0.0386	0.0143	0.0039	0.0007	0.0001	0.0000
	6	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128	0.4731	0.3348	0.2127	0.1178	0.0544	0.0194	0.0046	0.0005	0.0000
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062	0.6956	0.5618	0.4167	0.2763	0.1576	0.0726	0.0239	0.0043	0.0002
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270	0.8655	0.7747	0.6533	0.5075	0.3512	0.2054	0.0922	0.0256	0.0022
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807	0.9579	0.9166	0.8487	0.7472	0.6093	0.4417	0.2642	0.1109	00.196
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968	0.9917	0.9804	0.9576	0.9150	0.8416	0.7251	0.5565	0.3410	0.1184
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9992	0.9978	0.9943	0.9862	0.9683	0.9313	0.8578	0.7176	0.4596
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	0	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.8646	0.6213	0.3983	0.2336	0.1267	0.0637	0.0296	0.0126	0.0049	0.0017	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9755	0.8661	0.6920	0.5017	0.3326	0.2025	0.1132	0.0579	0.0269	0.0112	0.0041	0.0013	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9969	0.9658	0.8820	0.7473	0.5843	0.4206	0.2783	0.1686	0.0929	0.0461	0.0203	0.0078	0.0025	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	4	0.9997	0.9935	0.9658	0.9009	0.7940	0.6543	0.5005	0.3530	0.2279	0.1334	0.0698	0.0321	0.0126	0.0040	0.0010	0.0002	0.0000	0.0000	0.0000
	5	1.0000	0.9991	0.9925	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905	0.1788	0.0977	0.0462	0.0182	0.0056	0.0012	0.0002	0.0000	0.0000
	6	1.0000	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000	0.3563	0.2288	0.1295	0.0624	0.0243	0.0070	0.0013	0.0001	0.0000
	7	1.0000	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095	0.5732	0.4256	0.2841	0.1654	0.0802	0.0300	0.0075	0.0009	0.0000
	8	1.0000	1.0000	1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666	0.7721	0.6470	0.4995	0.3457	0.2060	0.0991	0.0342	0.0065	0.0003
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539	0.9071	0.8314	0.7217	0.5794	0.4157	0.2527	0.1180	0.0342	0.0031
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9959	0.9888	0.9731	0.9421	0.8868	0.7975	0.6674	0.4983	0.3080	0.1339	0.0245
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9951	0.9874	0.9704	0.9363	0.8733	0.7664	0.6017	0.3787	0.1354
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9963	0.9903	0.9762	0.9450	0.8791	0.7458	0.4867
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.8470	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9699	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065	0.0022	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9958	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287	0.0114	0.0039	0.0011	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.9996	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898	0.0426	0.0175	0.0060	0.0017	0.0003	0.0000	0.0000	0.0000	0.0000
	5	1.0000	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120	0.1189	0.0583	0.0243	0.0083	0.0022	0.0004	0.0000	0.0000	0.0000
	6	1.0000	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953	0.2586	0.1501	0.0753	0.0315	0.0103	0.0024	0.0003	0.0000	0.0000
	7	1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047	0.4539	0.3075	0.1836	0.0933	0.0383	0.0116	0.0022	0.0002	0.0000
	8	1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880	0.6627	0.5141	0.3595	0.2195	0.1117	0.0439	0.0115	0.0015	0.0000
	9	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102	0.8328	0.7207	0.5773	0.4158	0.2585	0.1298	0.0467	0.0092	0.0004
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713	0.9368	0.8757	0.7795	0.6448	0.4787	0.3018	0.1465	0.0441	0.0042
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935	0.9830	0.9602	0.9161	0.8392	0.7189	0.5519	0.3521	0.1584	0.0301
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991	0.9971	0.9919	0.9795	0.9525	0.8990	0.8021	0.6433	0.4154	0.1530
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9992	0.9976	0.9932	0.9822	0.9560	0.8972	0.7712	0.5123
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

（续）

n	y	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
15	0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176	0.0063	0.0019	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592	0.0255	0.0093	0.0028	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509	0.0769	0.0338	0.0124	0.0037	0.0008	0.0001	0.0000	0.0000	0.0000
	6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036	0.1818	0.0950	0.0422	0.0152	0.0042	0.0008	0.0001	0.0000	0.0000
	7	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000	0.3465	0.2131	0.1132	0.0500	0.0173	0.0042	0.0006	0.0000	0.0000
	8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964	0.5478	0.3902	0.2452	0.1311	0.0566	0.0181	0.0036	0.0003	0.0000
	9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491	0.7392	0.5968	0.4357	0.2784	0.1484	0.0611	0.0168	0.0022	0.0001
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408	0.8796	0.7827	0.6481	0.4845	0.3135	0.1642	0.0617	0.0127	0.0006
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824	0.9576	0.9095	0.8273	0.7031	0.5387	0.3518	0.1773	0.0556	0.0055
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963	0.9893	0.9729	0.9383	0.8732	0.7639	0.6020	0.3958	0.1841	0.0362
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9948	0.9858	0.9647	0.9198	0.8329	0.6814	0.4510	0.1710
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9984	0.9953	0.9866	0.9648	0.9126	0.7941	0.5367
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	0	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.8108	0.5147	0.2839	0.1407	0.0635	0.0261	0.0098	0.0033	0.0010	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9571	0.7892	0.5614	0.3518	0.1971	0.0994	0.0451	0.0183	0.0066	0.0021	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9930	0.9316	0.7899	0.5981	0.4050	0.2459	0.1339	0.0651	0.0281	0.0106	0.0035	0.0009	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.9991	0.9830	0.9209	0.7982	0.6302	0.4499	0.2892	0.1666	0.0853	0.0384	0.0149	0.0049	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.9999	0.9967	0.9765	0.9183	0.8103	0.6598	0.4900	0.3288	0.1976	0.1051	0.0486	0.0191	0.0062	0.0016	0.0003	0.0000	0.0000	0.0000	0.0000
	6	1.0000	0.9995	0.9944	0.9733	0.9204	0.8247	0.6881	0.5272	0.3660	0.2272	0.1241	0.0583	0.0229	0.0071	0.0016	0.0002	0.0000	0.0000	0.0000
	7	1.0000	0.9999	0.9989	0.9930	0.9729	0.9256	0.8406	0.7161	0.5629	0.4018	0.2559	0.1423	0.0671	0.0257	0.0075	0.0015	0.0002	0.0000	0.0000
	8	1.0000	1.0000	0.9998	0.9985	0.9925	0.9743	0.9329	0.8577	0.7441	0.5982	0.4371	0.2839	0.1594	0.0744	0.0271	0.0070	0.0011	0.0001	0.0000
	9	1.0000	1.0000	1.0000	0.9998	0.9984	0.9929	0.9771	0.9417	0.8759	0.7728	0.6340	0.4728	0.3119	0.1753	0.0796	0.0267	0.0056	0.0005	0.0000
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9938	0.9809	0.9514	0.8949	0.8024	0.6712	0.5100	0.3402	0.1897	0.0817	0.0235	0.0033	0.0001
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9851	0.9616	0.9147	0.8334	0.7108	0.5501	0.3698	0.2018	0.0791	0.0170	0.0009
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9965	0.9894	0.9719	0.9349	0.8661	0.7541	0.5950	0.4019	0.2101	0.0684	0.0070
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9979	0.9934	0.9817	0.9549	0.9006	0.8029	0.6482	0.4386	0.2108	0.0429
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9967	0.9902	0.9739	0.9365	0.8593	0.7161	0.4853	0.1892
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9967	0.9900	0.9719	0.9257	0.8147	0.5599
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

（续）

n	y	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
17	0	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.7922	0.4818	0.2525	0.1182	0.0501	0.0193	0.0067	0.0021	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9497	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9912	0.9174	0.7556	0.5489	0.3530	0.2019	0.1028	0.0464	0.0184	0.0064	0.0019	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.9988	0.9779	0.9013	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245	0.0086	0.0025	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.9999	0.9953	0.9681	0.8943	0.7653	0.5968	0.4197	0.2639	0.1471	0.0717	0.0301	0.0106	0.0030	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	6	1.0000	0.9992	0.9917	0.9623	0.8929	0.7752	0.6188	0.4478	0.2902	0.1662	0.0826	0.0348	0.0120	0.0032	0.0006	0.0001	0.0000	0.0000	0.0000
	7	1.0000	0.9999	0.9983	0.9891	0.9598	0.8954	0.7872	0.6405	0.4743	0.3145	0.1834	0.0919	0.0383	0.0127	0.0031	0.0005	0.0000	0.0000	0.0000
	8	1.0000	1.0000	0.9997	0.9974	0.9876	0.9597	0.9006	0.8011	0.6626	0.5000	0.3374	0.1989	0.0994	0.0403	0.0124	0.0026	0.0003	0.0000	0.0000
	9	1.0000	1.0000	1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6855	0.5257	0.3595	0.2128	0.1046	0.0402	0.0109	0.0017	0.0001	0.0000
	10	1.0000	1.0000	1.0000	0.9999	0.9994	0.9968	0.9880	0.9652	0.9174	0.8338	0.7098	0.5522	0.3812	0.2248	0.1071	0.0377	0.0083	0.0008	0.0000
	11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9970	0.9894	0.9699	0.9283	0.8529	0.7361	0.5803	0.4032	0.2347	0.1057	0.0319	0.0047	0.0001
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9975	0.9914	0.9755	0.9404	0.8740	0.7652	0.6113	0.4261	0.2418	0.0987	0.0221	0.0012
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9936	0.9816	0.9536	0.8972	0.7981	0.6470	0.4511	0.2444	0.0826	0.0088
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9988	0.9959	0.9877	0.9673	0.9226	0.8363	0.6904	0.4802	0.2382	0.0503
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9979	0.9933	0.9807	0.9499	0.8818	0.7475	0.5182	0.2078
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9977	0.9925	0.9775	0.9369	0.8332	0.5819
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	0.3972	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.7735	0.4503	0.2241	0.0991	0.0395	0.0142	0.0046	0.0013	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9419	0.7338	0.4797	0.2713	0.1353	0.0600	0.0236	0.0082	0.0025	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9891	0.9018	0.7202	0.5010	0.3057	0.1646	0.0783	0.0328	0.0120	0.0038	0.0010	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.9985	0.9718	0.8794	0.7164	0.5187	0.3327	0.1886	0.0942	0.0411	0.0154	0.0049	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.9998	0.9936	0.9581	0.8671	0.7175	0.5344	0.3550	0.2088	0.1077	0.0481	0.0183	0.0058	0.0014	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	6	1.0000	0.9988	0.9882	0.9487	0.8610	0.7217	0.5491	0.3743	0.2258	0.1189	0.0537	0.0203	0.0062	0.0014	0.0002	0.0000	0.0000	0.0000	0.0000
	7	1.0000	0.9998	0.9973	0.9837	0.9431	0.8593	0.7283	0.5634	0.3915	0.2403	0.1280	0.0576	0.0212	0.0061	0.0012	0.0002	0.0000	0.0000	0.0000
	8	1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7368	0.5778	0.4073	0.2527	0.1347	0.0597	0.0210	0.0054	0.0009	0.0001	0.0000	0.0000
	9	1.0000	1.0000	0.9999	0.9991	0.9946	0.9790	0.9403	0.8653	0.7473	0.5927	0.4222	0.2632	0.1391	0.0596	0.0193	0.0043	0.0005	0.0000	0.0000
	10	1.0000	1.0000	1.0000	0.9998	0.9988	0.9939	0.9788	0.9424	0.8720	0.7597	0.6085	0.4366	0.2717	0.1407	0.0569	0.0163	0.0027	0.0002	0.0000
	11	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9938	0.9797	0.9463	0.8811	0.7742	0.6257	0.4509	0.2783	0.1390	0.0513	0.0118	0.0012	0.0000
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9986	0.9942	0.9817	0.9519	0.8923	0.7912	0.6450	0.4656	0.2825	0.1329	0.0419	0.0064	0.0002
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9846	0.9589	0.9058	0.8114	0.6673	0.4813	0.2836	0.1206	0.0282	0.0015
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9962	0.9880	0.9672	0.9217	0.8354	0.6943	0.4990	0.2798	0.0982	0.0109
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9918	0.9764	0.9400	0.8647	0.7287	0.5203	0.2662	0.0581
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9954	0.9858	0.9605	0.9009	0.7759	0.5497	0.2265
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9944	0.9820	0.9464	0.8499	0.6028
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

（续）

n	y	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
19	0	0.3774	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.7547	0.4203	0.1985	0.0829	0.0310	0.0104	0.0031	0.0008	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9335	0.7054	0.4413	0.2369	0.1113	0.0462	0.0170	0.0055	0.0015	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9868	0.8850	0.6841	0.4551	0.2631	0.1332	0.0591	0.0230	0.0077	0.0022	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.9980	0.9648	0.8556	0.6733	0.4654	0.2822	0.1500	0.0696	0.0280	0.0096	0.0028	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.9998	0.9914	0.9463	0.8369	0.6678	0.4739	0.2968	0.1629	0.0777	0.0318	0.0109	0.0031	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	6	1.0000	0.9983	0.9837	0.9324	0.8251	0.6655	0.4812	0.3081	0.1727	0.0835	0.0342	0.0116	0.0031	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000
	7	1.0000	0.9997	0.9959	0.9767	0.9225	0.8180	0.6656	0.4878	0.3169	0.1796	0.0871	0.0352	0.0114	0.0028	0.0005	0.0000	0.0000	0.0000	0.0000
	8	1.0000	1.0000	0.9992	0.9933	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238	0.1841	0.0885	0.0347	0.0105	0.0023	0.0003	0.0000	0.0000	0.0000
	9	1.0000	1.0000	0.9999	0.9984	0.9911	0.9674	0.9125	0.8139	0.6710	0.5000	0.3290	0.1861	0.0875	0.0326	0.0089	0.0016	0.0001	0.0000	0.0000
	10	1.0000	1.0000	1.0000	0.9997	0.9977	0.9895	0.9653	0.9115	0.8159	0.6762	0.5060	0.3325	0.1855	0.0839	0.0287	0.0067	0.0008	0.0000	0.0000
	11	1.0000	1.0000	1.0000	1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204	0.6831	0.5122	0.3344	0.1820	0.0775	0.0233	0.0041	0.0003	0.0000
	12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9884	0.9658	0.9165	0.8273	0.6919	0.5188	0.3345	0.1749	0.0676	0.0163	0.0017	0.0000
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9969	0.9891	0.9682	0.9223	0.8371	0.7032	0.5261	0.3322	0.1631	0.0537	0.0086	0.0002
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9904	0.9720	0.9304	0.8500	0.7178	0.5346	0.3267	0.1444	0.0352	0.0020
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9978	0.9923	0.9770	0.9409	0.8668	0.7369	0.5449	0.3159	0.1150	0.0132
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9985	0.9945	0.9830	0.9538	0.8887	0.7631	0.5587	0.2946	0.0665
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9969	0.9896	0.9690	0.9171	0.8015	0.5797	0.2453
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9958	0.9856	0.9544	0.8649	0.6226
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059	0.0015	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207	0.0064	0.0016	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577	0.0214	0.0065	0.0015	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316	0.0580	0.0210	0.0060	0.0013	0.0002	0.0000	0.0000	0.0000	0.0000
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517	0.1308	0.0565	0.0196	0.0051	0.0009	0.0001	0.0000	0.0000	0.0000
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119	0.2493	0.1275	0.0532	0.0171	0.0039	0.0006	0.0000	0.0000	0.0000
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881	0.4086	0.2447	0.1218	0.0480	0.0139	0.0026	0.0002	0.0000	0.0000
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483	0.5857	0.4044	0.2376	0.1133	0.0409	0.0100	0.0013	0.0001	0.0000
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684	0.7480	0.5841	0.3990	0.2277	0.1018	0.0321	0.0059	0.0004	0.0000
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423	0.8701	0.7500	0.5834	0.3920	0.2142	0.0867	0.0219	0.0024	0.0000
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793	0.9447	0.8744	0.7546	0.5836	0.3828	0.1958	0.0673	0.0113	0.0003
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941	0.9811	0.9490	0.8818	0.7625	0.5852	0.3704	0.1702	0.0432	0.0026
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9840	0.9556	0.8929	0.7748	0.5886	0.3523	0.1330	0.0159
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9964	0.9879	0.9645	0.9087	0.7939	0.5951	0.3231	0.0755
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9979	0.9924	0.9757	0.9308	0.8244	0.6083	0.2642
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9968	0.9885	0.9612	0.8784	0.6415
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

a Y服从参数为n和p的二项分布. 表中的数是$P(Y \leq y) = \sum_{i=0}^{y} \binom{n}{i} p^i (1-p)^{n-i}$的值，$p$的范围从 0.05 到 0.95.

对于 n 大于20，其二项分布r分位数 y_r 可以近似使用 $y_r = np + z_r \sqrt{np(1-p)}$ 得到，这里 z_r 是表A1中标准正态随机变量的第 r 分位数.

表 A4　二项参数 p 的精确置信区间

n	Y	90% 左端点	90% 右端点	95% 左端点	95% 右端点	99% 左端点	99% 右端点
1	0	0.000	0.950	0.000	0.975	0.000	0.995
	1	0.050	1.000	0.025	1.000	0.005	1.000
2	0	0.000	0.776	0.000	0.842	0.000	0.929
	1	0.025	0.975	0.013	0.987	0.003	0.997
	2	0.224	1.000	0.158	1.000	0.071	1.000
3	0	0.000	0.632	0.000	0.708	0.000	0.829
	1	0.017	0.865	0.008	0.906	0.002	0.959
	2	0.135	0.983	0.094	0.992	0.041	0.998
	3	0.368	1.000	0.292	1.000	0.171	1.000
4	0	0.000	0.527	0.000	0.602	0.000	0.734
	1	0.013	0.751	0.006	0.806	0.001	0.889
	2	0.098	0.902	0.068	0.932	0.029	0.971
	3	0.249	0.987	0.194	0.994	0.111	0.999
	4	0.473	1.000	0.398	1.000	0.266	1.000
5	0	0.000	0.451	0.000	0.522	0.000	0.653
	1	0.010	0.657	0.005	0.716	0.001	0.815
	2	0.076	0.811	0.053	0.853	0.023	0.917
	3	0.189	0.924	0.147	0.947	0.083	0.977
	4	0.343	0.990	0.284	0.995	0.185	0.999
	5	0.549	1.000	0.478	1.000	0.347	1.000
6	0	0.000	0.393	0.000	0.459	0.000	0.586
	1	0.009	0.582	0.004	0.641	0.001	0.746
	2	0.063	0.729	0.043	0.777	0.019	0.856
	3	0.153	0.847	0.118	0.882	0.066	0.934
	4	0.271	0.937	0.223	0.957	0.144	0.981
	5	0.418	0.991	0.359	0.996	0.254	0.999
	6	0.607	1.000	0.541	1.000	0.414	1.000
7	0	0.000	0.348	0.000	0.410	0.000	0.531
	1	0.007	0.521	0.004	0.579	0.001	0.685
	2	0.053	0.659	0.037	0.710	0.016	0.797
	3	0.129	0.775	0.099	0.816	0.055	0.882
	4	0.225	0.871	0.184	0.901	0.118	0.945
	5	0.341	0.947	0.290	0.963	0.203	0.984
	6	0.479	0.993	0.421	0.996	0.315	0.999
	7	0.652	1.000	0.590	1.000	0.469	1.000
8	0	0.000	0.312	0.000	0.369	0.000	0.484
	1	0.006	0.471	0.003	0.526	0.001	0.632
	2	0.046	0.600	0.032	0.651	0.014	0.742
	3	0.111	0.711	0.085	0.755	0.047	0.830
	4	0.193	0.807	0.157	0.843	0.100	0.900
	5	0.289	0.889	0.245	0.915	0.170	0.953
	6	0.400	0.954	0.349	0.968	0.258	0.986
	7	0.529	0.994	0.474	0.997	0.368	0.999
	8	0.688	1.000	0.631	1.000	0.516	1.000

（续）

n	Y	90%		95%		99%	
		左端点	右端点	左端点	右端点	左端点	右端点
9	0	0.000	0.283	0.000	0.336	0.000	0.445
	1	0.006	0.429	0.003	0.482	0.001	0.585
	2	0.041	0.550	0.028	0.600	0.012	0.693
	3	0.098	0.655	0.075	0.701	0.042	0.781
	4	0.169	0.749	0.137	0.788	0.087	0.854
	5	0.251	0.831	0.212	0.863	0.146	0.913
	6	0.345	0.902	0.299	0.925	0.219	0.958
	7	0.450	0.959	0.400	0.972	0.307	0.988
	8	0.571	0.994	0.518	0.997	0.415	0.999
	9	0.717	1.000	0.664	1.000	0.555	1.000
10	0	0.000	0.259	0.000	0.308	0.000	0.411
	1	0.005	0.394	0.003	0.445	0.001	0.544
	2	0.037	0.507	0.025	0.556	0.011	0.648
	3	0.087	0.607	0.067	0.652	0.037	0.735
	4	0.150	0.696	0.122	0.738	0.077	0.809
	5	0.222	0.778	0.187	0.813	0.128	0.872
	6	0.304	0.850	0.262	0.878	0.191	0.923
	7	0.393	0.913	0.348	0.933	0.265	0.963
	8	0.493	0.963	0.444	0.975	0.352	0.989
	9	0.606	0.995	0.555	0.997	0.456	0.999
	10	0.741	1.000	0.692	1.000	0.589	1.000
11	0	0.000	0.238	0.000	0.285	0.000	0.382
	1	0.005	0.364	0.002	0.413	0.000	0.509
	2	0.033	0.470	0.023	0.518	0.010	0.608
	3	0.079	0.564	0.060	0.610	0.033	0.693
	4	0.135	0.650	0.109	0.692	0.069	0.767
	5	0.200	0.729	0.167	0.766	0.115	0.831
	6	0.271	0.800	0.234	0.833	0.169	0.885
	7	0.350	0.865	0.308	0.891	0.233	0.931
	8	0.436	0.921	0.390	0.940	0.307	0.967
	9	0.530	0.967	0.482	0.977	0.392	0.990
	10	0.636	0.995	0.587	0.998	0.491	1.000
	11	0.762	1.000	0.715	1.000	0.618	1.000
12	0	0.000	0.221	0.000	0.265	0.000	0.357
	1	0.004	0.339	0.002	0.385	0.000	0.477
	2	0.030	0.438	0.021	0.484	0.009	0.573
	3	0.072	0.527	0.055	0.572	0.030	0.655
	4	0.123	0.609	0.099	0.651	0.062	0.728
	5	0.181	0.685	0.152	0.723	0.103	0.792
	6	0.245	0.755	0.211	0.789	0.152	0.848
	7	0.315	0.819	0.277	0.848	0.208	0.897
	8	0.391	0.877	0.349	0.901	0.272	0.938
	9	0.473	0.928	0.428	0.945	0.345	0.970
	10	0.562	0.970	0.516	0.979	0.427	0.991
	11	0.661	0.996	0.615	0.998	0.523	1.000
	12	0.779	1.000	0.735	1.000	0.643	1.000

（续）

		90%		95%		99%	
n	*Y*	左端点	右端点	左端点	右端点	左端点	右端点
13	0	0.000	0.206	0.000	0.247	0.000	0.335
	1	0.004	0.316	0.002	0.360	0.000	0.449
	2	0.028	0.410	0.019	0.454	0.008	0.541
	3	0.066	0.495	0.050	0.538	0.028	0.621
	4	0.113	0.573	0.091	0.614	0.057	0.691
	5	0.166	0.645	0.139	0.684	0.094	0.755
	6	0.224	0.713	0.192	0.749	0.138	0.811
	7	0.287	0.776	0.251	0.808	0.189	0.862
	8	0.355	0.834	0.316	0.861	0.245	0.906
	9	0.427	0.887	0.386	0.909	0.309	0.943
	10	0.505	0.934	0.462	0.950	0.379	0.972
	11	0.590	0.972	0.546	0.981	0.459	0.992
	12	0.684	0.996	0.640	0.998	0.551	1.000
	13	0.794	1.000	0.753	1.000	0.665	1.000
14	0	0.000	0.193	0.000	0.232	0.000	0.315
	1	0.004	0.297	0.002	0.339	0.000	0.424
	2	0.026	0.385	0.018	0.428	0.008	0.512
	3	0.061	0.466	0.047	0.508	0.026	0.589
	4	0.104	0.540	0.084	0.581	0.053	0.658
	5	0.153	0.610	0.128	0.649	0.087	0.720
	6	0.206	0.675	0.177	0.711	0.127	0.777
	7	0.264	0.736	0.230	0.770	0.172	0.828
	8	0.325	0.794	0.289	0.823	0.223	0.873
	9	0.390	0.847	0.351	0.872	0.280	0.913
	10	0.460	0.896	0.419	0.916	0.342	0.947
	11	0.534	0.939	0.492	0.953	0.411	0.974
	12	0.615	0.974	0.572	0.982	0.488	0.992
	13	0.703	0.996	0.661	0.998	0.576	1.000
	14	0.807	1.000	0.768	1.000	0.685	1.000
15	0	0.000	0.181	0.000	0.218	0.000	0.298
	1	0.003	0.279	0.002	0.319	0.000	0.402
	2	0.024	0.363	0.017	0.405	0.007	0.486
	3	0.057	0.440	0.043	0.481	0.024	0.561
	4	0.097	0.511	0.078	0.551	0.049	0.627
	5	0.142	0.577	0.118	0.616	0.080	0.688
	6	0.191	0.640	0.163	0.677	0.117	0.744
	7	0.244	0.700	0.213	0.734	0.159	0.795
	8	0.300	0.756	0.266	0.787	0.205	0.841
	9	0.360	0.809	0.323	0.837	0.256	0.883
	10	0.423	0.858	0.384	0.882	0.312	0.920
	11	0.489	0.903	0.449	0.922	0.373	0.951
	12	0.560	0.943	0.519	0.957	0.439	0.976
	13	0.637	0.976	0.595	0.983	0.514	0.993
	14	0.721	0.997	0.681	0.998	0.598	1.000
	15	0.819	1.000	0.782	1.000	0.702	1.000
16	0	0.000	0.171	0.000	0.206	0.000	0.282
	1	0.003	0.264	0.002	0.302	0.000	0.381

（续）

n	Y	90% 左端点	90% 右端点	95% 左端点	95% 右端点	99% 左端点	99% 右端点
	2	0.023	0.344	0.016	0.383	0.007	0.463
	3	0.053	0.417	0.040	0.456	0.022	0.534
	4	0.090	0.484	0.073	0.524	0.045	0.599
	5	0.132	0.548	0.110	0.587	0.075	0.658
	6	0.178	0.609	0.152	0.646	0.109	0.713
	7	0.227	0.667	0.198	0.701	0.147	0.764
	8	0.279	0.721	0.247	0.753	0.190	0.810
	9	0.333	0.773	0.299	0.802	0.236	0.853
	10	0.391	0.822	0.354	0.848	0.287	0.891
	11	0.452	0.868	0.413	0.890	0.342	0.925
	12	0.516	0.910	0.476	0.927	0.401	0.955
	13	0.583	0.947	0.544	0.960	0.466	0.978
	14	0.656	0.977	0.617	0.984	0.537	0.993
	15	0.736	0.997	0.698	0.998	0.619	1.000
	16	0.829	1.000	0.794	1.000	0.718	1.000
17	0	0.000	0.162	0.000	0.195	0.000	0.268
	1	0.003	0.250	0.001	0.287	0.000	0.363
	2	0.021	0.326	0.015	0.364	0.006	0.441
	3	0.050	0.396	0.038	0.434	0.021	0.510
	4	0.085	0.461	0.068	0.499	0.043	0.573
	5	0.124	0.522	0.103	0.560	0.070	0.631
	6	0.166	0.580	0.142	0.617	0.101	0.685
	7	0.212	0.636	0.184	0.671	0.137	0.734
	8	0.260	0.689	0.230	0.722	0.176	0.781
	9	0.311	0.740	0.278	0.770	0.219	0.824
	10	0.364	0.788	0.329	0.816	0.266	0.863
	11	0.420	0.834	0.383	0.858	0.315	0.899
	12	0.478	0.876	0.440	0.897	0.369	0.930
	13	0.539	0.915	0.501	0.932	0.427	0.957
	14	0.604	0.950	0.566	0.962	0.490	0.979
	15	0.674	0.979	0.636	0.985	0.559	0.994
	16	0.750	0.997	0.713	0.999	0.637	1.000
	17	0.838	1.000	0.805	1.000	0.732	1.000
18	0	0.000	0.153	0.000	0.185	0.000	0.255
	1	0.003	0.238	0.001	0.273	0.000	0.346
	2	0.020	0.310	0.014	0.347	0.006	0.422
	3	0.047	0.377	0.036	0.414	0.020	0.488
	4	0.080	0.439	0.064	0.476	0.040	0.549
	5	0.116	0.498	0.097	0.535	0.065	0.605
	6	0.156	0.554	0.133	0.590	0.095	0.658
	7	0.199	0.608	0.173	0.643	0.128	0.707
	8	0.244	0.659	0.215	0.692	0.165	0.753
	9	0.291	0.709	0.260	0.740	0.205	0.795
	10	0.341	0.756	0.308	0.785	0.247	0.835
	11	0.392	0.801	0.357	0.827	0.293	0.872
	12	0.446	0.844	0.410	0.867	0.342	0.905
	13	0.502	0.884	0.465	0.903	0.395	0.935

（续）

n	Y	90%		95%		99%	
		左端点	右端点	左端点	右端点	左端点	右端点
	14	0.561	0.920	0.524	0.936	0.451	0.960
	15	0.623	0.953	0.586	0.964	0.512	0.980
	16	0.690	0.980	0.653	0.986	0.578	0.994
	17	0.762	0.997	0.727	0.999	0.654	1.000
	18	0.847	1.000	0.815	1.000	0.745	1.000
19	0	0.000	0.146	0.000	0.176	0.000	0.243
	1	0.003	0.226	0.001	0.260	0.000	0.331
	2	0.019	0.296	0.013	0.331	0.006	0.404
	3	0.044	0.359	0.034	0.396	0.019	0.468
	4	0.075	0.419	0.061	0.456	0.038	0.527
	5	0.110	0.476	0.091	0.512	0.062	0.582
	6	0.147	0.530	0.126	0.565	0.089	0.633
	7	0.188	0.582	0.163	0.616	0.121	0.681
	8	0.230	0.632	0.203	0.665	0.155	0.726
	9	0.274	0.680	0.244	0.711	0.192	0.768
	10	0.320	0.726	0.289	0.756	0.232	0.808
	11	0.368	0.770	0.335	0.797	0.274	0.845
	12	0.418	0.813	0.384	0.837	0.319	0.879
	13	0.470	0.853	0.435	0.874	0.367	0.911
	14	0.524	0.890	0.488	0.909	0.418	0.938
	15	0.581	0.925	0.544	0.939	0.473	0.962
	16	0.641	0.956	0.604	0.966	0.532	0.981
	17	0.704	0.981	0.669	0.987	0.596	0.994
	18	0.774	0.997	0.740	0.999	0.669	1.000
	19	0.854	1.000	0.824	1.000	0.757	1.000
20	0	0.000	0.139	0.000	0.168	0.000	0.233
	1	0.003	0.216	0.001	0.249	0.000	0.317
	2	0.018	0.283	0.012	0.317	0.005	0.387
	3	0.042	0.344	0.032	0.379	0.018	0.449
	4	0.071	0.401	0.057	0.437	0.036	0.507
	5	0.104	0.456	0.087	0.491	0.058	0.560
	6	0.140	0.508	0.119	0.543	0.085	0.610
	7	0.177	0.558	0.154	0.592	0.114	0.657
	8	0.217	0.606	0.191	0.639	0.146	0.701
	9	0.259	0.653	0.231	0.685	0.181	0.743
	10	0.302	0.698	0.272	0.728	0.218	0.782
	11	0.347	0.741	0.315	0.769	0.257	0.819
	12	0.394	0.783	0.361	0.809	0.299	0.854
	13	0.442	0.823	0.408	0.846	0.343	0.886
	14	0.492	0.860	0.457	0.881	0.390	0.915
	15	0.544	0.896	0.509	0.913	0.440	0.942
	16	0.599	0.929	0.563	0.943	0.493	0.964
	17	0.656	0.958	0.621	0.968	0.551	0.982
	18	0.717	0.982	0.683	0.988	0.613	0.995
	19	0.784	0.997	0.751	0.999	0.683	1.000
	20	0.861	1.000	0.832	1.000	0.767	1.000

（续）

n	Y	90%		95%		99%	
		左端点	右端点	左端点	右端点	左端点	右端点
21	0	0.000	0.133	0.000	0.161	0.000	0.223
	1	0.002	0.207	0.001	0.238	0.000	0.304
	2	0.017	0.271	0.012	0.304	0.005	0.372
	3	0.040	0.329	0.030	0.363	0.017	0.432
	4	0.068	0.384	0.054	0.419	0.034	0.488
	5	0.099	0.437	0.082	0.472	0.055	0.539
	6	0.132	0.487	0.113	0.522	0.080	0.588
	7	0.168	0.536	0.146	0.570	0.108	0.634
	8	0.206	0.583	0.181	0.616	0.138	0.677
	9	0.245	0.628	0.218	0.660	0.171	0.719
	10	0.286	0.672	0.257	0.702	0.205	0.758
	11	0.328	0.714	0.298	0.743	0.242	0.795
	12	0.372	0.755	0.340	0.782	0.281	0.829
	13	0.417	0.794	0.384	0.819	0.323	0.862
	14	0.464	0.832	0.430	0.854	0.366	0.892
	15	0.513	0.868	0.478	0.887	0.412	0.920
	16	0.563	0.901	0.528	0.918	0.461	0.945
	17	0.616	0.932	0.581	0.946	0.512	0.966
	18	0.671	0.960	0.637	0.970	0.568	0.983
	19	0.729	0.983	0.696	0.988	0.628	0.995
	20	0.793	0.998	0.762	0.999	0.696	1.000
	21	0.867	1.000	0.839	1.000	0.777	1.000
22	0	0.000	0.127	0.000	0.154	0.000	0.214
	1	0.002	0.198	0.001	0.228	0.000	0.292
	2	0.016	0.259	0.011	0.292	0.005	0.358
	3	0.038	0.316	0.029	0.349	0.016	0.416
	4	0.065	0.369	0.052	0.403	0.032	0.470
	5	0.094	0.420	0.078	0.454	0.053	0.520
	6	0.126	0.468	0.107	0.502	0.076	0.567
	7	0.160	0.515	0.139	0.549	0.102	0.612
	8	0.196	0.561	0.172	0.593	0.131	0.655
	9	0.233	0.605	0.207	0.636	0.162	0.695
	10	0.271	0.647	0.244	0.678	0.195	0.734
	11	0.311	0.689	0.282	0.718	0.229	0.771
	12	0.353	0.729	0.322	0.756	0.266	0.805
	13	0.395	0.767	0.364	0.793	0.305	0.838
	14	0.439	0.804	0.407	0.828	0.345	0.869
	15	0.485	0.840	0.451	0.861	0.388	0.898
	16	0.532	0.874	0.498	0.893	0.433	0.924
	17	0.580	0.906	0.546	0.922	0.480	0.947
	18	0.631	0.935	0.597	0.948	0.530	0.968
	19	0.684	0.962	0.651	0.971	0.584	0.984
	20	0.741	0.984	0.708	0.989	0.642	0.995

（续）

n	Y	90% 左端点	90% 右端点	95% 左端点	95% 右端点	99% 左端点	99% 右端点
	21	0.802	0.998	0.772	0.999	0.708	1.000
	22	0.873	1.000	0.846	1.000	0.786	1.000
23	0	0.000	0.122	0.000	0.148	0.000	0.206
	1	0.002	0.190	0.001	0.219	0.000	0.281
	2	0.016	0.249	0.011	0.280	0.005	0.345
	3	0.037	0.304	0.028	0.336	0.015	0.401
	4	0.062	0.355	0.050	0.388	0.031	0.453
	5	0.090	0.404	0.075	0.437	0.050	0.502
	6	0.120	0.451	0.102	0.484	0.073	0.548
	7	0.152	0.496	0.132	0.529	0.097	0.592
	8	0.186	0.540	0.164	0.573	0.125	0.634
	9	0.222	0.583	0.197	0.615	0.154	0.674
	10	0.258	0.625	0.232	0.655	0.185	0.712
	11	0.296	0.665	0.268	0.694	0.218	0.748
	12	0.335	0.704	0.306	0.732	0.252	0.782
	13	0.375	0.742	0.345	0.768	0.288	0.815
	14	0.417	0.778	0.385	0.803	0.326	0.846
	15	0.460	0.814	0.427	0.836	0.366	0.875
	16	0.504	0.848	0.471	0.868	0.408	0.903
	17	0.549	0.880	0.516	0.898	0.452	0.927
	18	0.596	0.910	0.563	0.925	0.498	0.950
	19	0.645	0.938	0.612	0.950	0.547	0.969
	20	0.696	0.963	0.664	0.972	0.599	0.985
	21	0.751	0.984	0.720	0.989	0.655	0.995
	22	0.810	0.998	0.781	0.999	0.719	1.000
	23	0.878	1.000	0.852	1.000	0.794	1.000
24	0	0.000	0.117	0.000	0.142	0.000	0.198
	1	0.002	0.183	0.001	0.211	0.000	0.271
	2	0.015	0.240	0.010	0.270	0.004	0.332
	3	0.035	0.292	0.027	0.324	0.015	0.387
	4	0.059	0.342	0.047	0.374	0.029	0.438
	5	0.086	0.389	0.071	0.422	0.048	0.485
	6	0.115	0.435	0.098	0.467	0.069	0.530
	7	0.146	0.479	0.126	0.511	0.093	0.573
	8	0.178	0.521	0.156	0.553	0.119	0.614
	9	0.212	0.563	0.188	0.594	0.146	0.653
	10	0.246	0.603	0.221	0.634	0.176	0.690
	11	0.282	0.642	0.256	0.672	0.207	0.726
	12	0.319	0.681	0.291	0.709	0.240	0.760
	13	0.358	0.718	0.328	0.744	0.274	0.793
	14	0.397	0.754	0.366	0.779	0.310	0.824
	15	0.437	0.788	0.406	0.812	0.347	0.854
	16	0.479	0.822	0.447	0.844	0.386	0.881
	17	0.521	0.854	0.489	0.874	0.427	0.907
	18	0.565	0.885	0.533	0.902	0.470	0.931
	19	0.611	0.914	0.578	0.929	0.515	0.952

（续）

n	Y	90% 左端点	90% 右端点	95% 左端点	95% 右端点	99% 左端点	99% 右端点
	20	0.658	0.941	0.626	0.953	0.562	0.971
	21	0.708	0.965	0.676	0.973	0.613	0.985
	22	0.760	0.985	0.730	0.990	0.668	0.996
	23	0.817	0.998	0.789	0.999	0.729	1.000
	24	0.883	1.000	0.858	1.000	0.802	1.000
25	0	0.000	0.113	0.000	0.137	0.000	0.191
	1	0.002	0.176	0.001	0.204	0.000	0.262
	2	0.014	0.231	0.010	0.260	0.004	0.321
	3	0.034	0.282	0.025	0.312	0.014	0.374
	4	0.057	0.330	0.045	0.361	0.028	0.424
	5	0.082	0.375	0.068	0.407	0.046	0.470
	6	0.110	0.420	0.094	0.451	0.066	0.514
	7	0.139	0.462	0.121	0.494	0.089	0.555
	8	0.170	0.504	0.150	0.535	0.114	0.595
	9	0.202	0.544	0.180	0.575	0.140	0.634
	10	0.236	0.583	0.211	0.613	0.168	0.670
	11	0.270	0.621	0.244	0.651	0.197	0.705
	12	0.305	0.659	0.278	0.687	0.228	0.739
	13	0.341	0.695	0.313	0.722	0.261	0.772
	14	0.379	0.730	0.349	0.756	0.295	0.803
	15	0.417	0.764	0.387	0.789	0.330	0.832
	16	0.456	0.798	0.425	0.820	0.366	0.860
	17	0.496	0.830	0.465	0.850	0.405	0.886
	18	0.538	0.861	0.506	0.879	0.445	0.911
	19	0.580	0.890	0.549	0.906	0.486	0.934
	20	0.625	0.918	0.593	0.932	0.530	0.954
	21	0.670	0.943	0.639	0.955	0.576	0.972
	22	0.718	0.966	0.688	0.975	0.626	0.986
	23	0.769	0.986	0.740	0.990	0.679	0.996
	24	0.824	0.998	0.796	0.999	0.738	1.000
	25	0.887	1.000	0.863	1.000	0.809	1.000
26	0	0.000	0.109	0.000	0.132	0.000	0.184
	1	0.002	0.170	0.001	0.196	0.000	0.253
	2	0.014	0.223	0.009	0.251	0.004	0.310
	3	0.032	0.272	0.024	0.302	0.013	0.362
	4	0.054	0.318	0.044	0.349	0.027	0.410
	5	0.079	0.363	0.066	0.393	0.044	0.455
	6	0.106	0.405	0.090	0.436	0.064	0.498
	7	0.134	0.447	0.116	0.478	0.085	0.538
	8	0.163	0.487	0.143	0.518	0.109	0.578
	9	0.194	0.526	0.172	0.557	0.134	0.615
	10	0.226	0.564	0.202	0.594	0.161	0.651
	11	0.258	0.602	0.234	0.631	0.189	0.686
	12	0.292	0.638	0.266	0.666	0.218	0.719
	13	0.327	0.673	0.299	0.701	0.249	0.751
	14	0.362	0.708	0.334	0.734	0.281	0.782

（续）

n	Y	90% 左端点	90% 右端点	95% 左端点	95% 右端点	99% 左端点	99% 右端点
	15	0.398	0.742	0.369	0.766	0.314	0.811
	16	0.436	0.774	0.406	0.798	0.349	0.839
	17	0.474	0.806	0.443	0.828	0.385	0.866
	18	0.513	0.837	0.482	0.857	0.422	0.891
	19	0.553	0.866	0.522	0.884	0.462	0.915
	20	0.595	0.894	0.564	0.910	0.502	0.936
	21	0.637	0.921	0.607	0.934	0.545	0.956
	22	0.682	0.946	0.651	0.956	0.590	0.973
	23	0.728	0.968	0.698	0.976	0.638	0.987
	24	0.777	0.986	0.749	0.991	0.690	0.996
	25	0.830	0.998	0.804	0.999	0.747	1.000
	26	0.891	1.000	0.868	1.000	0.816	1.000
27	0	0.000	0.105	0.000	0.128	0.000	0.178
	1	0.002	0.164	0.001	0.190	0.000	0.245
	2	0.013	0.215	0.009	0.243	0.004	0.300
	3	0.031	0.263	0.024	0.292	0.013	0.351
	4	0.052	0.308	0.042	0.337	0.026	0.397
	5	0.076	0.351	0.063	0.381	0.042	0.441
	6	0.101	0.392	0.086	0.423	0.061	0.483
	7	0.129	0.432	0.111	0.463	0.082	0.523
	8	0.157	0.471	0.138	0.502	0.104	0.561
	9	0.186	0.509	0.165	0.540	0.128	0.597
	10	0.217	0.547	0.194	0.576	0.154	0.633
	11	0.248	0.583	0.224	0.612	0.181	0.667
	12	0.280	0.618	0.255	0.647	0.209	0.700
	13	0.313	0.653	0.287	0.681	0.238	0.731
	14	0.347	0.687	0.319	0.713	0.269	0.762
	15	0.382	0.720	0.353	0.745	0.300	0.791
	16	0.417	0.752	0.388	0.776	0.333	0.819
	17	0.453	0.783	0.424	0.806	0.367	0.846
	18	0.491	0.814	0.460	0.835	0.403	0.872
	19	0.529	0.843	0.498	0.862	0.439	0.896
	20	0.568	0.871	0.537	0.889	0.477	0.918
	21	0.608	0.899	0.577	0.914	0.517	0.939
	22	0.649	0.924	0.619	0.937	0.559	0.958
	23	0.692	0.948	0.663	0.958	0.603	0.974
	24	0.737	0.969	0.708	0.976	0.649	0.987
	25	0.785	0.987	0.757	0.991	0.700	0.996
	26	0.836	0.998	0.810	0.999	0.755	1.000
	27	0.895	1.000	0.872	1.000	0.822	1.000
28	0	0.000	0.101	0.000	0.123	0.000	0.172
	1	0.002	0.159	0.001	0.183	0.000	0.237
	2	0.013	0.208	0.009	0.235	0.004	0.291
	3	0.030	0.254	0.023	0.282	0.012	0.340
	4	0.050	0.298	0.040	0.327	0.025	0.385

（续）

n	Y	90%		95%		99%	
		左端点	右端点	左端点	右端点	左端点	右端点
	5	0.073	0.339	0.061	0.369	0.041	0.428
	6	0.098	0.380	0.083	0.410	0.059	0.469
	7	0.124	0.419	0.107	0.449	0.079	0.508
	8	0.151	0.457	0.132	0.487	0.100	0.545
	9	0.179	0.494	0.159	0.524	0.123	0.581
	10	0.208	0.530	0.186	0.559	0.148	0.615
	11	0.238	0.565	0.215	0.594	0.173	0.649
	12	0.269	0.600	0.245	0.628	0.200	0.681
	13	0.301	0.634	0.275	0.661	0.228	0.713
	14	0.333	0.667	0.306	0.694	0.257	0.743
	15	0.366	0.699	0.339	0.725	0.287	0.772
	16	0.400	0.731	0.372	0.755	0.319	0.800
	17	0.435	0.762	0.406	0.785	0.351	0.827
	18	0.470	0.792	0.441	0.814	0.385	0.852
	19	0.506	0.821	0.476	0.841	0.419	0.877
	20	0.543	0.849	0.513	0.868	0.455	0.900
	21	0.581	0.876	0.551	0.893	0.492	0.921
	22	0.620	0.902	0.590	0.917	0.531	0.941
	23	0.661	0.927	0.631	0.939	0.572	0.959
	24	0.702	0.950	0.673	0.960	0.615	0.975
	25	0.746	0.970	0.718	0.977	0.660	0.988
	26	0.792	0.987	0.765	0.991	0.709	0.996
	27	0.841	0.998	0.817	0.999	0.763	1.000
	28	0.899	1.000	0.877	1.000	0.828	1.000
29	0	0.000	0.098	0.000	0.119	0.000	0.167
	1	0.002	0.153	0.001	0.178	0.000	0.230
	2	0.012	0.202	0.008	0.228	0.004	0.282
	3	0.029	0.246	0.022	0.274	0.012	0.330
	4	0.049	0.288	0.039	0.317	0.024	0.374
	5	0.070	0.329	0.058	0.358	0.039	0.416
	6	0.094	0.368	0.080	0.397	0.056	0.455
	7	0.119	0.406	0.103	0.435	0.076	0.493
	8	0.145	0.443	0.127	0.472	0.096	0.530
	9	0.172	0.479	0.153	0.508	0.119	0.565
	10	0.201	0.514	0.179	0.543	0.142	0.599
	11	0.229	0.549	0.207	0.577	0.167	0.632
	12	0.259	0.583	0.235	0.611	0.192	0.664
	13	0.289	0.616	0.264	0.643	0.219	0.695
	14	0.320	0.648	0.294	0.675	0.247	0.724
	15	0.352	0.680	0.325	0.706	0.276	0.753
	16	0.384	0.711	0.357	0.736	0.305	0.781
	17	0.417	0.741	0.389	0.765	0.336	0.808
	18	0.451	0.771	0.423	0.793	0.368	0.833
	19	0.486	0.799	0.457	0.821	0.401	0.858
	20	0.521	0.828	0.492	0.847	0.435	0.881
	21	0.557	0.855	0.528	0.873	0.470	0.904

（续）

n	Y	90%		95%		99%	
		左端点	右端点	左端点	右端点	左端点	右端点
	22	0.594	0.881	0.565	0.897	0.507	0.924
	23	0.632	0.906	0.603	0.920	0.545	0.944
	24	0.671	0.930	0.642	0.942	0.584	0.961
	25	0.712	0.951	0.683	0.961	0.626	0.976
	26	0.754	0.971	0.726	0.978	0.670	0.988
	27	0.798	0.988	0.772	0.992	0.718	0.996
	28	0.847	0.998	0.822	0.999	0.770	1.000
	29	0.902	1.000	0.881	1.000	0.833	1.000
30	0	0.000	0.095	0.000	0.116	0.000	0.162
	1	0.002	0.149	0.001	0.172	0.000	0.223
	2	0.012	0.195	0.008	0.221	0.004	0.274
	3	0.028	0.239	0.021	0.265	0.012	0.320
	4	0.047	0.280	0.038	0.307	0.023	0.363
	5	0.068	0.319	0.056	0.347	0.038	0.404
	6	0.091	0.357	0.077	0.386	0.054	0.443
	7	0.115	0.394	0.099	0.423	0.073	0.480
	8	0.140	0.430	0.123	0.459	0.093	0.516
	9	0.166	0.465	0.147	0.494	0.114	0.550
	10	0.193	0.499	0.173	0.528	0.137	0.583
	11	0.221	0.533	0.199	0.561	0.160	0.616
	12	0.250	0.566	0.227	0.594	0.185	0.647
	13	0.279	0.598	0.255	0.626	0.211	0.677
	14	0.308	0.630	0.283	0.657	0.237	0.707
	15	0.339	0.661	0.313	0.687	0.265	0.735
	16	0.370	0.692	0.343	0.717	0.293	0.763
	17	0.402	0.721	0.374	0.745	0.323	0.789
	18	0.434	0.750	0.406	0.773	0.353	0.815
	19	0.467	0.779	0.439	0.801	0.384	0.840
	20	0.501	0.807	0.472	0.827	0.417	0.863
	21	0.535	0.834	0.506	0.853	0.450	0.886
	22	0.570	0.860	0.541	0.877	0.484	0.907
	23	0.606	0.885	0.577	0.901	0.520	0.927
	24	0.643	0.909	0.614	0.923	0.557	0.946
	25	0.681	0.932	0.653	0.944	0.596	0.962
	26	0.720	0.953	0.693	0.962	0.637	0.977
	27	0.761	0.972	0.735	0.979	0.680	0.988
	28	0.805	0.988	0.779	0.992	0.726	0.996
	29	0.851	0.998	0.828	0.999	0.777	1.000
	30	0.905	1.000	0.884	1.000	0.838	1.000

来源：由 R. L. Iman 所作，并经允许使用.

表 A5 当 $r+m=1$ 时，非参数容忍限的样本容量[a]

$1-\alpha$	$q=0.500$	0.700	0.750	0.800	0.850	0.900	0.950	0.975	0.980	0.990
0.500	1	2	3	4	5	7	14	28	35	69
0.700	2	4	5	6	8	12	24	48	60	120
0.750	2	4	5	7	9	14	28	55	69	138
0.800	3	5	6	8	10	16	32	64	80	161
0.850	3	6	7	9	12	19	37	75	94	189
0.900	4	7	9	11	15	22	45	91	144	230
0.950	5	9	11	14	19	29	59	119	149	299
0.975	6	11	13	17	23	36	72	146	183	368
0.980	6	11	14	18	25	38	77	155	194	390
0.990	7	13	17	21	29	44	90	182	228	459
0.995	8	15	19	24	33	51	104	210	263	528
0.999	10	20	25	31	43	66	135	273	342	688

a 表中的数为样本容量 n，使得不等式 $q^n \leqslant \alpha$ 成立，并用于如 3.3 节中描述的求容忍限：

$P(X^{(1)} \leqslant$ 总体的 p 分位数$) \geqslant 1-\alpha$ 或

$P($总体的q分位数$\leqslant X^{(n)}) \geqslant 1-\alpha$.

表 A6 当 $r+m=2$ 时，非参数容忍限的样本容量[a]

$1-\alpha$	$q=0.500$	0.700	0.750	0.800	0.850	0.900	0.950	0.975	0.980	0.990
0.500	3	6	7	9	11	17	34	67	84	168
0.700	5	8	10	12	16	24	49	97	122	244
0.750	5	9	10	13	18	27	53	107	134	269
0.800	5	9	11	14	19	29	59	119	149	299
0.850	6	10	13	16	22	33	67	134	168	337
0.900	7	12	15	18	25	38	77	155	194	388
0.950	8	14	18	22	30	46	93	188	236	473
0.975	9	17	20	26	35	54	110	221	277	555
0.980	9	17	21	27	37	56	115	231	290	581
0.990	11	20	24	31	42	64	130	263	330	662
0.995	12	22	27	34	47	72	146	294	369	740
0.999	14	27	33	42	58	89	181	366	458	920

a 表中的数为样本容量 n，使得不等式 $q^n + nq^{n-1}(1-q) \leqslant \alpha$ 成立，并用于求容忍限：

当 $r+m=2$ 时，$P(X^{(r)} \leqslant$ 总体的 q 分位数 $\leqslant X^{(n+1-m)}) \geqslant 1-\alpha$.

表A7　Mann–Whitney检验统计量的分位数[a]

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	0.001	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	0.005	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4
	0.01	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	5	5
	0.025	3	3	3	3	3	3	4	4	4	5	5	5	5	5	5	6	6	6	6
	0.05	3	3	3	4	4	4	5	5	5	5	6	6	7	7	7	7	8	8	8
	0.10	3	4	4	5	5	5	6	6	7	7	8	8	8	9	9	10	10	11	11
3	0.001	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7
	0.005	6	6	6	6	6	6	6	7	7	7	8	8	8	9	9	9	9	10	10
	0.01	6	6	6	6	6	7	7	8	8	8	9	9	9	10	10	11	11	11	12
	0.025	6	6	6	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	15
	0.05	6	7	7	8	9	9	10	11	11	12	12	13	14	14	15	16	16	17	18
	0.10	7	8	8	9	10	11	12	12	13	14	15	16	17	17	18	19	20	21	22
4	0.001	10	10	10	10	10	10	10	10	11	11	11	12	12	12	13	13	14	14	14
	0.005	10	10	10	10	11	11	12	12	13	13	14	14	15	16	16	17	17	18	19
	0.01	10	10	10	11	12	12	13	14	14	15	16	16	17	18	18	19	20	20	21
	0.025	10	10	11	12	13	14	15	15	16	17	18	19	20	21	22	22	23	24	25
	0.05	10	11	12	13	14	15	16	17	18	19	20	21	22	23	25	26	27	28	29
	0.10	11	12	14	15	16	17	18	20	21	22	23	24	26	27	28	29	31	32	33
5	0.001	15	15	15	15	15	15	16	17	17	18	18	19	19	20	21	21	22	23	23
	0.005	15	15	15	16	17	17	18	19	20	21	22	23	23	24	25	26	27	28	29
	0.01	15	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
	0.025	15	16	17	18	19	21	22	23	24	25	27	28	29	30	31	33	34	35	36
	0.05	16	17	18	20	21	22	24	25	27	28	29	31	32	34	35	36	38	39	41
	0.10	17	18	20	21	23	24	26	28	29	31	33	34	36	38	39	41	43	44	46
6	0.001	21	21	21	21	21	21	23	24	25	26	26	27	28	29	30	31	32	33	34
	0.005	21	21	22	23	24	25	26	27	28	29	31	32	33	34	35	37	38	39	40
	0.01	21	21	23	24	25	26	28	29	30	31	33	34	35	37	38	40	41	42	44
	0.025	21	23	24	25	27	28	30	32	33	35	36	38	39	41	43	44	46	47	49
	0.05	22	24	25	27	29	30	32	34	36	38	39	41	43	45	47	48	50	52	54
	0.10	23	25	27	29	31	33	35	37	39	41	43	45	47	49	51	53	56	58	60
7	0.001	28	28	28	28	29	30	31	32	34	35	36	37	38	39	40	42	43	44	45
	0.005	28	28	29	30	32	33	35	36	38	39	41	42	44	45	47	48	50	51	53
	0.01	28	29	30	32	33	35	36	38	40	41	43	45	46	48	50	52	53	55	57
	0.025	28	30	32	34	35	37	39	41	43	45	47	49	51	53	55	57	59	61	63
	0.05	29	31	33	35	37	40	42	44	46	48	50	53	55	57	59	62	64	66	68
	0.10	30	33	35	37	40	42	45	47	50	52	55	57	60	62	65	67	70	72	75
8	0.001	36	36	36	37	38	39	41	42	43	45	46	48	49	51	52	54	55	57	58
	0.005	36	36	38	39	41	43	44	46	48	50	52	54	55	57	59	61	63	65	67
	0.01	36	37	39	41	43	44	46	48	50	52	54	56	59	61	63	65	67	69	71
	0.025	37	39	41	43	45	47	50	52	54	56	59	61	63	66	68	71	73	75	78
	0.05	38	40	42	45	47	50	52	55	57	60	63	65	68	70	73	76	78	81	84
	0.10	39	42	44	47	50	53	56	59	61	64	67	70	73	76	79	82	85	88	91

(续)

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
9	0.001	45	45	45	47	48	49	51	53	54	56	58	60	61	63	65	67	69	71	72
	0.005	45	46	47	49	51	53	55	57	59	62	64	66	68	70	73	75	77	79	82
	0.01	45	47	49	51	53	55	57	60	62	64	67	69	72	74	77	79	82	84	86
	0.025	46	48	50	53	56	58	61	63	66	69	72	74	77	80	83	85	88	91	94
	0.05	47	50	52	55	58	61	64	67	70	73	76	79	82	85	88	91	94	97	100
	0.10	48	51	55	58	61	64	68	71	74	77	81	84	87	91	94	98	101	104	108
10	0.001	55	55	56	57	59	61	62	64	66	68	70	73	75	77	79	81	83	85	88
	0.005	55	56	58	60	62	65	67	69	72	74	77	80	82	85	87	90	93	95	98
	0.01	55	57	59	62	64	67	69	72	75	78	80	83	86	89	92	94	97	100	103
	0.025	56	59	61	64	67	70	73	76	79	82	85	89	92	95	98	101	104	108	111
	0.05	57	60	63	67	70	73	76	80	83	87	90	93	97	100	104	107	111	114	118
	0.10	59	62	66	69	73	77	80	84	88	92	95	99	103	107	110	114	118	122	126
11	0.001	66	66	67	69	71	73	75	77	79	82	84	87	89	91	94	96	99	101	104
	0.005	66	67	69	72	74	77	80	83	85	88	91	94	97	100	103	106	109	112	115
	0.01	66	68	71	74	76	79	82	85	89	92	95	98	101	104	108	111	114	117	120
	0.025	67	70	73	76	80	83	86	90	93	97	100	104	107	111	114	118	122	125	129
	0.05	68	72	75	79	83	86	90	94	98	101	105	109	113	117	121	124	128	132	136
	0.10	70	74	78	82	86	90	94	98	103	107	111	115	119	124	128	132	136	140	145
12	0.001	78	78	79	81	83	86	88	91	93	96	98	102	104	106	110	113	116	118	121
	0.005	78	80	82	85	88	91	94	97	100	103	106	110	113	116	120	123	126	130	133
	0.01	78	81	84	87	90	93	96	100	103	107	110	114	117	121	125	128	132	135	139
	0.025	80	83	86	90	93	97	101	105	108	112	116	120	124	128	132	136	140	144	148
	0.05	81	84	88	92	96	100	105	109	111	117	121	126	130	134	139	143	147	151	156
	0.10	83	87	91	96	100	105	109	114	118	123	128	132	137	142	146	151	156	160	165
13	0.001	91	91	93	95	97	100	103	106	109	112	115	118	121	124	127	130	134	137	140
	0.005	91	93	95	99	102	105	109	112	116	119	123	126	130	134	137	141	145	149	152
	0.01	92	94	97	101	104	108	112	115	119	123	127	131	135	139	143	147	151	155	159
	0.025	93	96	100	104	108	112	116	120	125	129	133	137	142	146	151	155	159	164	168
	0.05	94	98	102	107	111	116	120	125	129	134	139	143	148	153	157	162	167	172	176
	0.10	96	101	105	110	115	120	125	130	135	140	145	150	155	160	166	171	176	181	186
14	0.001	105	105	107	109	112	115	118	121	125	128	131	135	138	142	145	149	152	156	160
	0.005	105	107	110	113	117	121	124	128	132	136	140	144	148	152	156	160	164	169	173
	0.01	106	108	112	116	119	123	128	132	136	140	144	149	153	157	162	166	171	175	179
	0.025	107	111	115	119	123	128	132	137	142	146	151	156	161	165	170	175	180	184	189
	0.05	109	113	117	122	127	132	137	142	147	152	157	162	167	172	177	183	188	193	198
	0.10	110	116	121	126	131	137	142	147	153	158	164	169	175	180	186	191	197	203	208
15	0.001	120	120	122	125	128	133	135	138	142	145	149	153	157	161	164	168	172	176	180
	0.005	120	123	126	129	133	137	141	145	150	154	158	163	167	172	176	181	185	190	194
	0.01	121	124	128	132	136	140	145	149	154	158	163	168	172	177	182	187	191	196	201
	0.025	122	126	131	135	140	145	150	155	160	165	170	175	180	185	191	196	201	206	211
	0.05	124	128	133	139	144	149	154	160	165	171	176	182	187	193	198	204	209	215	221
	0.10	126	131	137	143	148	154	160	166	172	178	184	189	195	201	207	213	219	225	231
16	0.001	136	136	139	142	145	148	152	156	160	164	168	172	176	180	185	189	193	197	202
	0.005	136	139	142	146	150	155	159	164	168	173	178	182	187	192	197	202	207	211	216
	0.01	137	140	144	149	153	158	163	168	173	178	183	188	193	198	203	208	213	219	224
	0.025	138	143	148	152	158	163	168	174	179	184	190	196	201	207	212	218	223	229	235
	0.05	140	145	151	156	162	167	173	179	185	191	197	202	208	214	220	226	232	238	244
	0.10	142	148	154	160	166	173	179	185	191	198	204	211	217	223	230	236	243	249	256

(续)

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
17	0.001	153	154	156	159	163	167	171	175	179	183	188	192	197	201	206	211	215	220	224
	0.005	153	156	160	164	169	173	178	183	188	193	198	203	208	214	219	224	229	235	240
	0.01	154	158	162	167	172	177	182	187	192	198	203	209	214	220	225	231	236	242	247
	0.025	156	160	165	171	176	182	188	193	199	205	211	217	223	229	235	241	247	253	259
	0.05	157	163	169	174	180	187	193	199	205	211	218	224	231	237	243	250	256	263	269
	0.10	160	166	172	179	185	192	199	206	212	219	226	233	239	246	253	260	267	274	281
18	0.001	171	172	175	178	182	186	190	195	199	204	209	214	218	223	228	233	238	243	248
	0.005	171	174	178	183	188	193	198	203	209	214	219	225	230	236	242	247	253	259	264
	0.01	172	176	181	186	191	196	202	208	213	219	225	231	237	242	248	254	260	266	272
	0.025	174	179	184	190	196	202	208	214	220	227	233	239	246	252	258	265	271	278	284
	0.05	176	181	188	194	200	207	213	220	227	233	240	247	254	260	267	274	281	288	295
	0.10	178	185	192	199	206	213	220	227	234	241	249	256	263	270	278	285	292	300	307
19	0.001	190	191	194	198	202	206	211	216	220	225	231	236	241	246	251	257	262	268	273
	0.005	191	194	198	203	208	213	219	224	230	236	242	248	254	260	265	272	278	284	290
	0.01	192	195	200	206	211	217	223	229	235	241	247	254	260	266	273	279	285	292	298
	0.025	193	198	204	210	216	223	229	236	243	249	256	263	269	276	283	290	297	304	310
	0.05	195	201	208	214	221	228	235	242	249	256	263	271	278	285	292	300	307	314	321
	0.10	198	205	212	219	227	234	242	249	257	264	272	280	288	295	303	311	319	326	334
20	0.001	210	211	214	218	223	227	232	237	243	248	253	259	265	270	276	281	287	293	299
	0.005	211	214	219	224	229	235	241	247	253	259	265	271	278	284	290	297	303	310	316
	0.01	212	216	221	227	233	239	245	251	258	264	271	278	284	291	298	304	311	318	325
	0.025	213	219	225	231	238	245	251	259	266	273	280	287	294	301	309	316	323	330	338
	0.05	215	222	229	236	243	250	258	265	273	280	288	295	303	311	318	326	334	341	349
	0.10	218	226	233	241	249	257	265	273	281	289	297	305	313	321	330	338	346	354	362

对于超过20的 m 或 n ,Mann-Whitney检验统计量的 p 分位数 w_p 可由表达式 $w_p = n(N+1)/2 + z_p\sqrt{nm(N+1)/12}$ 近似得到. 这里 z_p 是标准正态随机变量的分位数，它可从表A1获得, $N = n + m$.

a 表中的数是Mann-Whitney检验统计量 T 的分位数 w_p ，对选择的 p 值，由(5.1.1)式给出，即满足 $P(T < w_p) \leqslant p$. 上侧分位数可由式

$$w_p = n(n + m + 1) - w_{1-p}$$

求得, 临界域对应着 T 值小于（或大于），但不等于不合适的 p 分位数.

表 A8 小样本 Kruskal – Wallis 检验统计量的分位数[a]

样本大小	$w_{0.90}$	$w_{0.95}$	$w_{0.99}$
2, 2, 2	3.7143	4.5714	4.5714
3, 2, 1	3.8571	4.2857	4.2857
3, 2, 2	4.4643	4.5000	5.3571
3, 3, 1	4.0000	4.5714	5.1429
3, 3, 2	4.2500	5.1389	6.2500
3, 3, 3	4.6000	5.0667	6.4889
4, 2, 1	4.0179	4.8214	4.8214
4, 2, 2	4.1667	5.1250	6.0000
4, 3, 1	3.8889	5.0000	5.8333
4, 3, 2	4.4444	5.4000	6.3000
4, 3, 3	4.7000	5.7273	6.7091
4, 4, 1	4.0667	4.8667	6.1667
4, 4, 2	4.4455	5.2364	6.8727
4, 4, 3	4.7730	5.5758	7.1364
4, 4, 4	4.5000	5.6538	7.5385
5, 2, 1	4.0500	4.4500	5.2500
5, 2, 2	4.2933	5.0400	6.1333
5, 3, 1	3.8400	4.8711	6.4000
5, 3, 2	4.4946	5.1055	6.8218
5, 3, 3	4.4121	5.5152	6.9818
5, 4, 1	3.9600	4.8600	6.8400
5, 4, 2	4.5182	5.2682	7.1182
5, 4, 3	4.5231	5.6308	7.3949
5, 4, 4	4.6187	5.6176	7.7440
5, 5, 1	4.0364	4.9091	6.8364
5, 5, 2	4.5077	5.2462	7.2692
5, 5, 3	4.5363	5.6264	7.5429
5, 5, 4	4.5200	5.6429	7.7914
5, 5, 5	4.5000	5.6600	7.9800

来源：经 *American Mathematical Society* 允许，并由 Iman，Quade 及 Alexander（1975）改编.

a 如果由（5.2.5）式给出的 Kruskal-Wallis 检验统计量超过表中给出的 $1-\alpha$ 值，则在水平 α 下可拒绝零假设.

表 A9　平方秩检验统计量的分位数[a]

n	p	m = 3	4	5	6	7	8	9	10
3	0.005	14	14	14	14	14	14	21	21
	0.01	14	14	14	14	21	21	26	26
	0.025	14	14	21	26	29	30	35	41
	0.05	21	21	26	30	38	42	49	54
	0.10	26	29	35	42	50	59	69	77
	0.90	65	90	117	149	182	221	260	305
	0.95	70	101	129	161	197	238	285	333
	0.975	77	110	138	170	213	257	308	362
	0.99	77	110	149	194	230	285	329	394
	0.995	77	110	149	194	245	302	346	413
4	0.005	30	30	30	39	39	46	50	54
	0.01	30	30	39	46	50	51	62	66
	0.025	30	39	50	54	63	71	78	90
	0.05	39	50	57	66	78	90	102	114
	0.10	50	62	71	85	99	114	130	149
	0.90	111	142	182	222	270	321	375	435
	0.95	119	154	197	246	294	350	413	476
	0.975	126	165	206	255	311	374	439	510
	0.99	126	174	219	270	334	401	470	545
	0.995	126	174	230	281	351	414	494	567
5	0.005	55	55	66	75	79	88	99	110
	0.01	55	66	75	82	90	103	115	127
	0.025	66	79	88	100	114	130	145	162
	0.05	75	88	103	120	135	155	175	195
	0.10	87	103	121	142	163	187	212	239
	0.90	169	214	264	319	379	445	514	591
	0.95	178	228	282	342	410	479	558	639
	0.975	183	235	297	363	433	508	592	680
	0.99	190	246	310	382	459	543	631	727
	0.995	190	255	319	391	478	559	654	754
6	0.005	91	104	115	124	136	152	167	182
	0.01	91	115	124	139	155	175	191	210
	0.025	115	130	143	164	184	208	231	255
	0.05	124	139	164	187	211	239	268	299
	0.10	136	163	187	215	247	280	315	352
	0.90	243	300	364	435	511	592	679	772
	0.95	255	319	386	463	545	634	730	831
	0.975	259	331	406	486	574	670	771	880
	0.99	271	339	424	511	607	706	817	935
	0.995	271	346	431	526	624	731	847	970

来源：由 R. L. Iman 所作的原始表改编，经允许使用.

a 表中的数是由（5.3.3）式给出的平方秩检验统计量 T 的有选择的 p 分位数 w_p，它满足 $P(T<w_p)\leqslant p$ 和 $P(T>w_p)\leqslant 1-p$. 临界域对应着 T 值小于（或大于），但不等于合适的 p 分位数.

（续）

n	p	m = 3	4	5	6	7	8	9	10
7	0.005	140	155	172	195	212	235	257	280
	0.01	155	172	191	212	236	260	287	315
	0.025	172	195	217	245	274	305	338	372
	0.05	188	212	240	274	308	344	384	425
	0.10	203	236	271	308	350	394	440	489
	0.90	335	407	487	572	665	764	871	984
	0.95	347	428	515	608	707	814	929	1051
	0.975	356	443	536	635	741	856	979	1108
	0.99	364	456	560	664	779	900	1032	1172
	0.995	371	467	571	683	803	929	1067	1212
8	0.005	204	236	260	284	311	340	368	401
	0.01	221	249	276	309	340	372	408	445
	0.025	249	276	311	345	384	425	468	513
	0.05	268	300	340	381	426	473	524	576
	0.10	285	329	374	423	476	531	590	652
	0.90	447	536	632	735	846	965	1091	1224
	0.95	464	560	664	776	896	1023	1159	1303
	0.975	476	579	689	807	935	1071	1215	1368
	0.99	485	599	716	840	980	1124	1277	1442
	0.995	492	604	731	863	1005	1156	1319	1489
9	0.005	304	325	361	393	429	466	508	549
	0.01	321	349	384	423	464	508	553	601
	0.025	342	380	423	469	517	570	624	682
	0.05	365	406	457	510	567	626	689	755
	0.10	390	444	501	561	625	694	766	843
	0.90	581	689	803	925	1056	1195	1343	1498
	0.95	601	717	840	972	1112	1261	1420	1587
	0.975	615	741	870	1009	1158	1317	1485	1662
	0.99	624	757	900	1049	1209	1377	1556	1745
	0.995	629	769	916	1073	1239	1417	1601	1798
10	0.005	406	448	486	526	573	620	672	725
	0.01	425	470	513	561	613	667	725	785
	0.025	457	505	560	616	677	741	808	879
	0.05	486	539	601	665	734	806	883	963
	0.10	514	580	649	724	801	885	972	1064
	0.90	742	866	1001	1144	1296	1457	1627	1806
	0.95	765	901	1045	1197	1360	1533	1715	1907
	0.975	778	925	1078	1241	1413	1596	1788	1991
	0.99	793	949	1113	1286	1470	1664	1869	2085
	0.995	798	961	1130	1314	1505	1708	1921	2145

对于超过 10 的 n 或 m，平方秩检验统计量的 p 分位数 w_p 可由

$$w_p = \frac{n(N+1)(2N+1)}{6} + z_p\sqrt{\frac{mn(N+1)(2N+1)(8N+11)}{180}}$$

近似得到，这里 z_p 是标准正态分布的 p 分位数，可从表 A1 获得，$N = n + m$.

表 A10　Spearman ρ 的分位数[a]

n	p = 0.900	0.950	0.975	0.990	0.995	0.999
4	0.8000	0.8000				
5	0.7000	0.8000	0.9000	0.9000		
6	0.6000	0.7714	0.8286	0.8857	0.9429	
7	0.5357	0.6786	0.7500	0.8571	0.8929	0.9643
8	0.5000	0.6190	0.7143	0.8095	0.8571	0.9286
9	0.4667	0.5833	0.6833	0.7667	0.8167	0.9000
10	0.4424	0.5515	0.6364	0.7333	0.7818	0.8667
11	0.4182	0.5273	0.6091	0.7000	0.7455	0.8364
12	0.3986	0.4965	0.5804	0.6713	0.7203	0.8112
13	0.3791	0.4780	0.5549	0.6429	0.6978	0.7857
14	0.3626	0.4593	0.5341	0.6220	0.6747	0.7670
15	0.3500	0.4429	0.5179	0.6000	0.6500	0.7464
16	0.3382	0.4265	0.5000	0.5794	0.6324	0.7265
17	0.3260	0.4118	0.4853	0.5637	0.6152	0.7083
18	0.3148	0.3994	0.4696	0.5480	0.5975	0.6904
19	0.3070	0.3895	0.4579	0.5333	0.5825	0.6737
20	0.2977	0.3789	0.4451	0.5203	0.5684	0.6586
21	0.2909	0.3688	0.4351	0.5078	0.5545	0.6455
22	0.2829	0.3597	0.4241	0.4963	0.5426	0.6318
23	0.2767	0.3518	0.4150	0.4852	0.5306	0.6186
24	0.2704	0.3435	0.4061	0.4748	0.5200	0.6070
25	0.2646	0.3362	0.3977	0.4654	0.5100	0.5962
26	0.2588	0.3299	0.3894	0.4564	0.5002	0.5856
27	0.2540	0.3236	0.3822	0.4481	0.4915	0.5757
28	0.2490	0.3175	0.3749	0.4401	0.4828	0.5660
29	0.2443	0.3113	0.3685	0.4320	0.4744	0.5567
30	0.2400	0.3059	0.3620	0.4251	0.4665	0.5479

对于 n 大于 30，ρ 的近似值可由 $w_p \cong \frac{z_p}{\sqrt{n-1}}$ 得到. 这里 z_p 是从表 A1 获得的标准正态随机变量的 p 分位数.

来源：在 *Biometrika* 委托允许下由 Glasser 和 Winter(1961)修改而来.

a 表中的数为当 Spearman 秩相关系数 ρ 作为检验统计量时的 p 分位数 w_p. 下侧分位数由式 $w_p = -w_{1-p}$ 得到，临界域对应着 ρ 值小于（或大于），但不等于合适的 p 分位数. 注意，ρ 的中位数是 0.

表 A11　Kendall 检验统计量 $T = N_c - N_d$ 和 Kendall τ 的分位数（Kendall τ 的分位数给在左侧），下侧分位数是上侧分位数的负数 $w_p = -w_{1-p}$.

n	p = 0.900	0.950	0.975	0.990	0.995
4	4 (0.6667)	4 (0.6667)	6 (1.0000)	6 (1.0000)	6 (1.0000)
5	6 (0.6000)	6 (0.6000)	8 (0.8000)	8 (0.8000)	10 (1.0000)
6	7 (0.4667)	9 (0.6000)	11 (0.7333)	11 (0.7333)	13 (0.8667)
7	9 (0.4286)	11 (0.5238)	13 (0.6190)	15 (0.7143)	17 (0.8095)
8	10 (0.3571)	14 (0.5000)	16 (0.5714)	18 (0.6429)	20 (0.7143)
9	12 (0.3333)	16 (0.4444)	18 (0.5000)	22 (0.6111)	24 (0.6667)
10	15 (0.3333)	19 (0.4222)	21 (0.4667)	25 (0.5556)	27 (0.6000)
11	17 (0.3091)	21 (0.3818)	25 (0.4545)	29 (0.5273)	31 (0.5636)
12	18 (0.2727)	24 (0.3636)	28 (0.4242)	34 (0.5152)	36 (0.5455)
13	22 (0.2821)	26 (0.3333)	32 (0.4103)	38 (0.4872)	42 (0.5285)
14	23 (0.2527)	31 (0.3407)	35 (0.3846)	41 (0.4505)	45 (0.4945)
15	27 (0.2571)	33 (0.3143)	39 (0.3714)	47 (0.4476)	51 (0.4857)
16	28 (0.2333)	36 (0.3000)	44 (0.3667)	50 (0.4167)	56 (0.4667)
17	32 (0.2353)	40 (0.2941)	48 (0.3529)	56 (0.4118)	62 (0.4559)
18	35 (0.2288)	43 (0.2810)	51 (0.3333)	61 (0.3987)	67 (0.4379)
19	37 (0.2164)	47 (0.2749)	55 (0.3216)	65 (0.3801)	73 (0.4269)
20	40 (0.2105)	50 (0.2632)	60 (0.3158)	70 (0.3684)	78 (0.4105)
21	42 (0.2000)	54 (0.2571)	64 (0.3048)	76 (0.3619)	84 (0.4000)
22	45 (0.1948)	59 (0.2554)	69 (0.2987)	81 (0.3506)	89 (0.3853)
23	49 (0.1937)	63 (0.2490)	73 (0.2885)	87 (0.3439)	97 (0.3834)
24	52 (0.1884)	66 (0.2391)	78 (0.2826)	92 (0.3333)	102 (0.3696)
25	56 (0.1867)	70 (0.2333)	84 (0.2800)	98 (0.3267)	108 (0.3600)
26	59 (0.1815)	75 (0.2308)	89 (0.2738)	105 (0.3231)	115 (0.3538)
27	61 (0.1738)	79 (0.2251)	93 (0.2650)	111 (0.3162)	123 (0.3504)
28	66 (0.1746)	84 (0.2222)	98 (0.2593)	116 (0.3069)	128 (0.3386)
29	68 (0.1675)	88 (0.2167)	104 (0.2562)	124 (0.3054)	136 (0.3350)
30	73 (0.1678)	93 (0.2138)	109 (0.2506)	129 (0.2966)	143 (0.3287)
31	75 (0.1613)	97 (0.2086)	115 (0.2473)	135 (0.2903)	149 (0.3204)
32	80 (0.1613)	102 (0.2056)	120 (0.2419)	142 (0.2863)	158 (0.3185)
33	84 (0.1591)	106 (0.2008)	126 (0.2386)	150 (0.2841)	164 (0.3106)
34	87 (0.1551)	111 (0.1979)	131 (0.2335)	155 (0.2763)	173 (0.3084)
35	91 (0.1529)	115 (0.1933)	137 (0.2303)	163 (0.2739)	179 (0.3008)
36	94 (0.1492)	120 (0.1905)	144 (0.2286)	170 (0.2698)	188 (0.2984)
37	98 (0.1471)	126 (0.1892)	150 (0.2252)	176 (0.2643)	198 (0.2943)
38	103 (0.1465)	131 (0.1863)	155 (0.2205)	183 (0.2603)	203 (0.2888)
39	107 (0.1444)	137 (0.1849)	161 (0.2173)	191 (0.2578)	211 (0.2848)
40	110 (0.1372)	142 (0.1821)	168 (0.2154)	198 (0.2538)	220 (0.2821)
41	114 (0.1390)	146 (0.1780)	174 (0.2122)	206 (0.2512)	228 (0.2780)
42	119 (0.1382)	151 (0.1754)	181 (0.2102)	213 (0.2474)	235 (0.2729)
43	123 (0.1362)	157 (0.1739)	187 (0.2071)	221 (0.2447)	245 (0.2713)
44	128 (0.1353)	162 (0.1712)	194 (0.2051)	228 (0.2410)	252 (0.2664)
45	132 (0.1333)	168 (0.1697)	200 (0.2020)	236 (0.2383)	262 (0.2646)
46	135 (0.1304)	173 (0.1671)	207 (0.2000)	245 (0.2367)	271 (0.2618)
47	141 (0.1304)	179 (0.1656)	213 (0.1970)	253 (0.2340)	279 (0.2581)
48	144 (0.1277)	186 (0.1649)	220 (0.1950)	260 (0.2305)	288 (0.2553)
49	150 (0.1276)	190 (0.1616)	228 (0.1939)	268 (0.2279)	296 (0.2517)
50	153 (0.1249)	197 (0.1608)	233 (0.1902)	277 (0.2261)	305 (0.2490)
51	159 (0.1247)	203 (0.1592)	241 (0.1890)	285 (0.2235)	315 (0.2471)
52	162 (0.1222)	208 (0.1569)	248 (0.1870)	294 (0.2217)	324 (0.2443)
53	168 (0.1219)	214 (0.1553)	256 (0.1858)	302 (0.2192)	334 (0.2424)
54	173 (0.1209)	221 (0.1544)	263 (0.1838)	311 (0.2173)	343 (0.2397)
55	177 (0.1192)	227 (0.1529)	269 (0.1811)	319 (0.2148)	353 (0.2377)
56	182 (0.1182)	232 (0.1506)	276 (0.1792)	328 (0.2130)	362 (0.2351)
57	186 (0.1165)	240 (0.1504)	284 (0.1779)	336 (0.2105)	372 (0.2331)
58	191 (0.1155)	245 (0.1482)	291 (0.1760)	345 (0.2087)	381 (0.2305)
59	197 (0.1151)	251 (0.1467)	299 (0.1748)	355 (0.2075)	391 (0.2285)
60	202 (0.1141)	258 (0.1458)	306 (0.1729)	364 (0.2056)	402 (0.2271)

对于 n 大于60，T 的近似 p 分位数可由 $w_p \cong z_p\sqrt{\frac{n(n-1)(2n+5)}{18}}$ 得到，这里 z_p 是表 A1 给出的标准正态分布的 p 分位数，τ 的近似分位数可由 $w_p \cong z_p\frac{\sqrt{2(2n+5)}}{3\sqrt{n(n-1)}}$ 得到.

临界域对应着 T 值小于（或大于），但不等于某合适的 p 分位数. 注意，T 的中位数是 0，τ 的近似分位数可由 T 的分位数除以 $n(n-1)/2$ 得到.

来源：经作者同意由 Best(1974)表 1 改编.

表 A12　Wilcoxon 符号秩检验统计量的分位数[a]

	$w_{0.005}$	$w_{0.01}$	$w_{0.025}$	$w_{0.05}$	$w_{0.10}$	$w_{0.20}$	$w_{0.30}$	$w_{0.40}$	$w_{0.50}$	$\frac{n(n+1)}{2}$
$n=4$	0	0	0	0	1	3	3	4	5	10
5	0	0	0	1	3	4	5	6	7.5	15
6	0	0	1	3	4	6	8	9	10.5	21
7	0	1	3	4	6	9	11	12	14	28
8	1	2	4	6	9	12	14	16	18	36
9	2	4	6	9	11	15	18	20	22.5	45
10	4	6	9	11	15	19	22	25	27.5	55
11	6	8	11	14	18	23	27	30	33	66
12	8	10	14	18	22	28	32	36	39	78
13	10	13	18	22	27	33	38	42	45.5	91
14	13	16	22	26	32	39	44	48	52.5	105
15	16	20	26	31	37	45	51	55	60	120
16	20	24	30	36	43	51	58	63	68	136
17	24	28	35	42	49	58	65	71	76.5	153
18	28	33	41	48	56	66	73	80	85.5	171
19	33	38	47	54	63	74	82	89	95	190
20	38	44	53	61	70	83	91	98	105	210
21	44	50	59	68	78	91	100	108	115.5	231
22	49	56	67	76	87	100	110	119	126.5	253
23	55	63	74	84	95	110	120	130	138	276
24	62	70	82	92	105	120	131	141	150	300
25	69	77	90	101	114	131	143	153	162.5	325
26	76	85	99	111	125	142	155	165	175.5	351
27	84	94	108	120	135	154	167	178	189	378
28	92	102	117	131	146	166	180	192	203	406
29	101	111	127	141	158	178	193	206	217.5	435
30	110	121	138	152	170	191	207	220	232.5	465
31	119	131	148	164	182	205	221	235	248	496
32	129	141	160	176	195	219	236	250	264	528
33	139	152	171	188	208	233	251	266	280.5	561
34	149	163	183	201	222	248	266	282	297.5	595
35	160	175	196	214	236	263	283	299	315	630
36	172	187	209	228	251	279	299	317	333	666
37	184	199	222	242	266	295	316	335	351.5	703
38	196	212	236	257	282	312	334	353	370.5	741
39	208	225	250	272	298	329	352	372	390	780
40	221	239	265	287	314	347	371	391	410	820
41	235	253	280	303	331	365	390	411	430.5	861
42	248	267	295	320	349	384	409	431	451.5	903
43	263	282	311	337	366	403	429	452	473	946
44	277	297	328	354	385	422	450	473	495	990
45	292	313	344	372	403	442	471	495	517.5	1035
46	308	329	362	390	423	463	492	517	540.5	1081
47	324	346	379	408	442	484	514	540	564	1128
48	340	363	397	428	463	505	536	563	588	1176
49	357	381	416	447	483	527	559	587	612.5	1225
50	374	398	435	467	504	550	583	611	637.5	1275

对于 n 超过 50，Wilcoxon 符号秩检验统计量的 p 分位数可由

$$w_p = [n(n+1)/4] + z_p\sqrt{n(n+1)(2n+1)/24}$$

近似得到. 这里 z_p 是标准正态分布的 p 分位数，可从表 A1 获得.

来源：由 Harter 和 Owen（1970）改编，经 *American Mathematical Society* 允许使用.

a 表中的数是 Wilcoxon 符号秩检验统计量 T^+ 的 p 分位数 w_p，选择 $p \leqslant 0.50$ 由（5.7.3）式给出. 对于 $p > 0.50$ 可由等式 $w_p = n(n+1)/2 - w_{1-p}$ 计算得到，这里 $n(n+1)/2$ 是表中最右边的列. 注意，如果 H_0 成立，则有 $P(T^+ < w_p) \leqslant p$ 和 $P(T^+ > w_p) \leqslant 1-p$. 临界域对应着 T^+ 值小于（或大于），但不等于某合适的 p 分位数.

表 A13 Kolmogorov 检验统计量的分位数[a]

单边检验	p = 0.90	0.95	0.975	0.99	0.995		p = 0.90	0.95	0.975	0.99	0.995
双边检验	p = 0.80	0.90	0.95	0.98	0.99		p = 0.80	0.90	0.95	0.98	0.99
n = 1	0.900	0.950	0.975	0.990	0.995	n = 21	0.226	0.259	0.287	0.321	0.344
2	0.684	0.776	0.842	0.900	0.929	22	0.221	0.253	0.281	0.314	0.337
3	0.565	0.636	0.708	0.785	0.829	23	0.216	0.247	0.275	0.307	0.330
4	0.493	0.565	0.624	0.689	0.734	24	0.212	0.242	0.269	0.301	0.323
5	0.447	0.509	0.563	0.627	0.669	25	0.208	0.238	0.264	0.295	0.317
6	0.410	0.468	0.519	0.577	0.617	26	0.204	0.233	0.259	0.290	0.311
7	0.381	0.436	0.483	0.538	0.576	27	0.200	0.229	0.254	0.284	0.305
8	0.358	0.410	0.454	0.507	0.542	28	0.197	0.225	0.250	0.279	0.300
9	0.339	0.387	0.430	0.480	0.513	29	0.193	0.221	0.246	0.275	0.295
10	0.323	0.369	0.409	0.457	0.489	30	0.190	0.218	0.242	0.270	0.290
11	0.308	0.352	0.391	0.437	0.468	31	0.187	0.214	0.238	0.266	0.285
12	0.296	0.338	0.375	0.419	0.449	32	0.184	0.211	0.234	0.262	0.281
13	0.285	0.325	0.361	0.404	0.432	33	0.182	0.208	0.231	0.258	0.277
14	0.275	0.314	0.349	0.390	0.418	34	0.179	0.205	0.227	0.254	0.273
15	0.266	0.304	0.338	0.377	0.404	35	0.177	0.202	0.224	0.251	0.269
16	0.258	0.295	0.327	0.366	0.392	36	0.174	0.199	0.221	0.247	0.265
17	0.250	0.286	0.318	0.355	0.381	37	0.172	0.196	0.218	0.244	0.262
18	0.244	0.279	0.309	0.346	0.371	38	0.170	0.194	0.215	0.241	0.258
19	0.237	0.271	0.301	0.337	0.361	39	0.168	0.191	0.213	0.238	0.255
20	0.232	0.265	0.294	0.329	0.352	40	0.165	0.189	0.210	0.235	0.252
					>40 的近似值		$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

来源：由 Miller（1956）改编，经 *American Statistical Association* 允许使用.

a 表中的数是由（6.1.1）式定义的 Kolmogorov 双边检验统计量 T，由（6.1.2）和（6.1.3）式定义的 T^- 和 T^- 单边检验的有选择的 p 分位数 w_p. 如果 T 超过表中 $1-\alpha$ 分位数，则以水平 α 拒绝 H_0. 在 $n \leqslant 40$ 的双边检验中，这些分位数是精确值，其余的分位数是近似值，但在大多数情形下等于精确值. 对于 $n>40$，分母则用 $(n+\sqrt{n/10})^{1/2}$ 代替$\sqrt{n}$，则近似会更好些.

表 A14　Lilliefors 正态性检验统计量的分位数[a]

		$p = 0.80$	0.85	0.90	0.95	0.99
样本大小	$n = 4$	0.303	0.320	0.344	0.374	0.414
	5	0.290	0.302	0.319	0.344	0.398
	6	0.268	0.280	0.295	0.321	0.371
	7	0.252	0.264	0.280	0.304	0.353
	8	0.239	0.251	0.266	0.290	0.333
	9	0.227	0.239	0.253	0.275	0.319
	10	0.217	0.228	0.241	0.262	0.303
	11	0.209	0.219	0.232	0.252	0.291
	12	0.201	0.210	0.223	0.243	0.281
	13	0.193	0.203	0.215	0.233	0.270
	14	0.187	0.196	0.209	0.227	0.264
	15	0.181	0.190	0.202	0.219	0.256
	16	0.176	0.184	0.195	0.212	0.248
	17	0.170	0.179	0.190	0.207	0.241
	18	0.166	0.174	0.185	0.201	0.234
	19	0.162	0.171	0.181	0.197	0.230
	20	0.159	0.167	0.177	0.192	0.223
	21	0.155	0.163	0.173	0.188	0.219
	22	0.152	0.160	0.170	0.185	0.214
	23	0.149	0.156	0.165	0.181	0.210
	24	0.145	0.153	0.162	0.177	0.205
	25	0.144	0.151	0.159	0.173	0.202
	26	0.141	0.147	0.156	0.170	0.198
	27	0.138	0.145	0.153	0.166	0.193
	28	0.136	0.142	0.151	0.165	0.191
	29	0.134	0.140	0.149	0.162	0.188
	30	0.132	0.138	0.146	0.159	0.183
	$\geqslant 31$	$\frac{0.741}{d_n}$	$\frac{0.775}{d_n}$	$\frac{0.819}{d_n}$	$\frac{0.895}{d_n}$	$\frac{1.035}{d_n}$

$d_n = (\sqrt{n} - 0.01 + 0.83/\sqrt{n})$

来源：Mason 和 Bell(1986)表 L.5，经 Marcel Dekker，Inc 允许使用.

a 表中的数是（6.2.4）式定义的 Lilliefors 检验统计量 T_1 的分位数 w_p 的近似值，对指定样本大小为 n，如果 T_1 超过 $w_{1-\alpha}$，则以水平 α 拒绝 H_0.

表A15 Lilliefors指数分布检验统计量的分位数[a]

	p = 0.05	0.10	0.20	0.30	0.50	0.70	0.80	0.90	0.95	0.99	0.999
n = 2	0.3127	0.3200	0.3337	0.3617	0.4337	0.5034	0.5507	0.5934	0.6133	0.6284	0.6317
3	0.2299	0.2544	0.2899	0.3166	0.3645	0.4122	0.4508	0.5111	0.5508	0.6003	0.6296
4	0.2072	0.2281	0.2545	0.2766	0.3163	0.3685	0.4007	0.4442	0.4844	0.5574	0.6215
5	0.1884	0.2052	0.2290	0.2483	0.2877	0.3317	0.3603	0.4045	0.4420	0.5127	0.5814
6	0.1726	0.1882	0.2102	0.2290	0.2645	0.3045	0.3320	0.3732	0.4085	0.4748	0.5497
7	0.1604	0.1750	0.1961	0.2136	0.2458	0.2838	0.3098	0.3481	0.3811	0.4459	0.5181
8	0.1506	0.1646	0.1845	0.2006	0.2309	0.2671	0.2914	0.3274	0.3590	0.4208	0.4913
9	0.1426	0.1561	0.1746	0.1897	0.2186	0.2529	0.2758	0.3101	0.3404	0.3995	0.4679
10	0.1359	0.1486	0.1661	0.1805	0.2082	0.2407	0.2626	0.2955	0.3244	0.3813	0.4473
12	0.1249	0.1364	0.1524	0.1657	0.1912	0.2209	0.2411	0.2714	0.2981	0.3511	0.4132
14	0.1162	0.1268	0.1418	0.1542	0.1778	0.2054	0.2242	0.2525	0.2774	0.3272	0.3858
16	0.1091	0.1191	0.1332	0.1448	0.1669	0.1929	0.2105	0.2371	0.2606	0.3076	0.3632
18	0.1032	0.1127	0.1260	0.1369	0.1578	0.1824	0.1990	0.2242	0.2465	0.2911	0.3441
20	0.0982	0.1073	0.1199	0.1303	0.1501	0.1735	0.1893	0.2132	0.2345	0.2271	0.3277
22	0.0939	0.1025	0.1146	0.1245	0.1434	0.1657	0.1809	0.2038	0.2241	0.2649	0.3135
24	0.0901	0.0984	0.1099	0.1195	0.1376	0.1590	0.1735	0.1954	0.2150	0.2542	0.3010
26	0.0868	0.0947	0.1058	0.1150	0.1324	0.1530	0.1670	0.1881	0.2069	0.2447	0.2899
28	0.0838	0.0914	0.1021	0.1110	0.1278	0.1477	0.1611	0.1815	0.1997	0.2362	0.2799
30	0.0811	0.0885	0.0988	0.1074	0.1236	0.1428	0.1559	0.1756	0.1932	0.2286	0.2709
35	0.0754	0.0822	0.0918	0.0997	0.1148	0.1326	0.1447	0.1630	0.1793	0.2123	0.2517
40	0.0707	0.0771	0.0861	0.0935	0.1077	0.1243	0.1356	0.1528	0.1681	0.1990	0.2361
45	0.0668	0.0729	0.0814	0.0884	0.1017	0.1174	0.1281	0.1443	0.1588	0.1880	0.2231
50	0.0636	0.0693	0.0774	0.0840	0.0966	0.1116	0.1217	0.1371	0.1509	0.1787	0.2121
60	0.0582	0.0635	0.0708	0.0769	0.0885	0.1021	0.1114	0.1255	0.1381	0.1635	0.1943
70	0.0541	0.0589	0.0658	0.0714	0.0821	0.0946	0.1033	0.1164	0.1281	0.1517	b
80	0.0507	0.0553	0.0616	0.0669	0.0769	0.0887	0.0968	0.1090	0.1200	0.1421	b
90	0.0479	0.0522	0.0582	0.0632	0.0726	0.0838	0.0914	0.1029	0.1132	0.1341	b
n = 100	0.0455	0.0496	0.0553	0.0600	0.0690	0.0796	0.0868	0.0977	0.1075	0.1274	b
n>100的近似值	$\frac{0.4550}{\sqrt{n}}$	$\frac{0.4959}{\sqrt{n}}$	$\frac{0.5530}{\sqrt{n}}$	$\frac{0.6000}{\sqrt{n}}$	$\frac{0.6898}{\sqrt{n}}$	$\frac{0.7957}{\sqrt{n}}$	$\frac{0.8678}{\sqrt{n}}$	$\frac{0.9773}{\sqrt{n}}$	$\frac{1.0753}{\sqrt{n}}$	$\frac{1.2743}{\sqrt{n}}$	b

来源：经Biometrika委托许可由Durbin (1975)修改而来.

a 表中的数是(6.2.6)式给出的Lilliefors检验统计量T_2的分位数w_p的值，如果T_2超过表中$1-\alpha$分位数，则以显著水平α拒绝H_0. 对于n>100时，近似值仅仅当n=100时是精确值，对n>100导出更精确的近似可以从Pearson和Hartley表54得到.

b 这些分位数没有在表中列出.

表 A16　Shapiro-Wilk 检验的系数[a]

i \ n	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.0588	0.5739
2	—	0.0000	0.1667	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291
3	—	—	—	0.0000	0.0875	0.1401	0.1743	0.1976	0.2141
4	—	—	—	—	—	0.0000	0.0561	0.0947	0.1224
5	—	—	—	—	—	—	—	0.0000	0.0399

i \ n	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211
3	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565
4	0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085
5	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7	—	—	0.0000	0.0240	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013
8	—	—	—	—	0.0000	0.0196	0.0359	0.0496	0.0612	0.0711
9	—	—	—	—	—	—	0.0000	0.0163	0.0303	0.0422
10	—	—	—	—	—	—	—	—	0.0000	0.0140

i \ n	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944
3	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499	0.2487
4	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148
5	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864	0.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1395	0.1415
8	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192	0.1219
9	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002	0.1036
10	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822	0.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12	—	—	0.0000	0.0107	0.0200	0.0284	0.0358	0.0424	0.0483	0.0537
13	—	—	—	—	0.0000	0.0094	0.0178	0.0253	0.0320	0.0381
14	—	—	—	—	—	—	0.0000	0.0084	0.0159	0.0227
15	—	—	—	—	—	—	—	—	0.0000	0.0076

（续）

i \ n	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	0.2921	0.2898	0.2876	0.2854	0.2834	0.2813	0.2794	0.2774	0.2755	0.2737
3	0.2475	0.2462	0.2451	0.2439	0.2427	0.2415	0.2403	0.2391	0.2380	0.2368
4	0.2145	0.2141	0.2137	0.2132	0.2127	0.2121	0.2116	0.2110	0.2104	0.2098
5	0.1874	0.1878	0.1880	0.1882	0.1883	0.1883	0.1883	0.1881	0.1880	0.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	0.1433	0.1449	0.1463	0.1475	0.1487	0.1496	0.1505	0.1513	0.1520	0.1526
8	0.1243	0.1265	0.1284	0.1301	0.1317	0.1331	0.1344	0.1356	0.1366	0.1376
9	0.1066	0.1093	0.1118	0.1140	0.1160	0.1179	0.1196	0.1211	0.1225	0.1237
10	0.0899	0.0931	0.0961	0.0988	0.1013	0.1036	0.1056	0.1075	0.1092	0.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	0.0585	0.0629	0.0669	0.0706	0.0739	0.0770	0.0798	0.0824	0.0848	0.0870
13	0.0435	0.0485	0.0530	0.0572	0.0610	0.0645	0.0677	0.0706	0.0733	0.0759
14	0.0289	0.0344	0.0395	0.0441	0.0484	0.0523	0.0559	0.0592	0.0622	0.0651
15	0.0144	0.0206	0.0262	0.0314	0.0361	0.0404	0.0444	0.0481	0.0515	0.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17	—	—	0.0000	0.0062	0.0119	0.0172	0.0220	0.0264	0.0305	0.0343
18	—	—	—	—	0.0000	0.0057	0.0110	0.0158	0.0203	0.0244
19	—	—	—	—	—	—	0.0000	0.0053	0.0101	0.0146
20	—	—	—	—	—	—	—	—	0.0000	0.0049

i \ n	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574
3	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260
4	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032
5	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
8	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
9	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
10	0.1123	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.0998	0.1010	0.1020
13	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.0919	0.0932
14	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
15	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
18	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
19	0.0188	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
20	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22	—	—	0.0000	0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
23	—	—	—	—	0.0000	0.0039	0.0076	0.0111	0.0143	0.0174
24	—	—	—	—	—	—	0.0000	0.0037	0.0071	0.0104
25	—	—	—	—	—	—	—	—	0.0000	0.0035

来源：经 *Biometrika* Trustees 许可，由 Pearson 和 Hartley（1976）第二卷再版修改而来.
a 表中的数是用由（6.2.9）式给出的 Shapiro-Wilk 正态性检验统计量中的系数 a_i.

表 A17　Shapiro-Wilk 检验统计量的分位数[a]

n	0.01	0.02	0.05	0.10	0.50	0.90	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	0.687	0.707	0.748	0.792	0.935	0.987	0.992	0.996	0.997
5	0.686	0.715	0.762	0.806	0.927	0.979	0.986	0.991	0.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	0.730	0.760	0.803	0.838	0.928	0.972	0.979	0.985	0.988
8	0.749	0.778	0.818	0.851	0.932	0.972	0.978	0.984	0.987
9	0.764	0.791	0.829	0.859	0.935	0.972	0.978	0.984	0.986
10	0.781	0.806	0.842	0.869	0.938	0.972	0.978	0.983	0.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	0.805	0.828	0.859	0.883	0.943	0.973	0.979	0.984	0.986
13	0.814	0.837	0.866	0.889	0.945	0.974	0.979	0.984	0.986
14	0.825	0.846	0.874	0.895	0.947	0.975	0.980	0.984	0.986
15	0.835	0.855	0.881	0.901	0.950	0.975	0.980	0.984	0.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	0.851	0.869	0.892	0.910	0.954	0.977	0.981	0.985	0.987
18	0.858	0.874	0.897	0.914	0.956	0.978	0.982	0.986	0.988
19	0.863	0.879	0.901	0.917	0.957	0.978	0.982	0.986	0.988
20	0.868	0.884	0.905	0.920	0.959	0.979	0.983	0.986	0.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	0.878	0.892	0.911	0.926	0.961	0.980	0.984	0.987	0.989
23	0.881	0.895	0.914	0.928	0.962	0.981	0.984	0.987	0.989
24	0.884	0.898	0.916	0.930	0.963	0.981	0.984	0.987	0.989
25	0.888	0.901	0.918	0.931	0.964	0.981	0.985	0.988	0.989
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989
27	0.894	0.906	0.923	0.935	0.965	0.982	0.985	0.988	0.990
28	0.896	0.908	0.924	0.936	0.966	0.982	0.985	0.988	0.990
29	0.898	0.910	0.926	0.937	0.966	0.982	0.985	0.988	0.990
30	0.900	0.912	0.927	0.939	0.967	0.983	0.985	0.988	0.990
31	0.902	0.914	0.929	0.940	0.967	0.983	0.986	0.988	0.990
32	0.904	0.915	0.930	0.941	0.968	0.983	0.986	0.988	0.990
33	0.906	0.917	0.931	0.942	0.968	0.983	0.986	0.989	0.990
34	0.908	0.919	0.933	0.943	0.969	0.983	0.986	0.989	0.990
35	0.910	0.920	0.934	0.944	0.969	0.984	0.986	0.989	0.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	0.914	0.924	0.936	0.946	0.970	0.984	0.987	0.989	0.990
38	0.916	0.925	0.938	0.947	0.971	0.984	0.987	0.989	0.990
39	0.917	0.927	0.939	0.948	0.971	0.984	0.987	0.989	0.991
40	0.919	0.928	0.940	0.949	0.972	0.985	0.987	0.989	0.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	0.922	0.930	0.942	0.951	0.972	0.985	0.987	0.989	0.991
43	0.923	0.932	0.943	0.951	0.973	0.985	0.987	0.990	0.991
44	0.924	0.933	0.944	0.952	0.973	0.985	0.987	0.990	0.991
45	0.926	0.934	0.945	0.953	0.973	0.985	0.988	0.990	0.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	0.928	0.936	0.946	0.954	0.974	0.985	0.988	0.990	0.991
48	0.929	0.937	0.947	0.954	0.974	0.985	0.988	0.990	0.991
49	0.929	0.937	0.947	0.955	0.974	0.985	0.988	0.990	0.991
50	0.930	0.938	0.947	0.955	0.974	0.985	0.988	0.990	0.991

来源：经 *Biometrika* 委托许可由 Pearson 和 Hartley（1976）再版修改而来.

a 表中的数是由（6.2.9）式给出的 Shapiro-Wilk 检验统计量的 p 分位数 w_p，如果 $T_3 < w_p$，则以水平 α 拒绝 H_0.

表 A18　变换 Shapiro-Wilk 统计量到近似正态分布的方法

$v \backslash n$ (d_n)	3 (0.7500)	4 (0.6297)	5 (0.5521)	6 (0.4963)	n v (d_n)	3 (0.7500)	4 (0.6297)	5 (0.5521)	6 (0.4963)
−7.0	−3.29	—	—	—	**2.2**	0.52	0.74	0.75	0.64
−5.4	−2.81	—	—	—	**2.6**	0.67	1.00	1.09	1.06
−5.0	−2.68	—	—	—	**3.0**	0.81	1.23	1.40	1.45
−4.6	−2.54	—	—	—	**3.4**	0.95	1.44	1.67	1.83
−4.2	−2.40	—	—	—	**3.8**	1.07	1.65	1.91	2.17
−3.8	−2.25	−3.50	—	—	**4.2**	1.19	1.85	2.15	2.50
−3.4	−2.10	−3.27	—	—	**4.6**	1.31	2.03	2.47	2.77
−3.0	−1.94	−3.05	−4.01	—	**5.0**	1.42	2.19	2.85	3.09
−2.6	−1.77	−2.84	−3.70	—	**5.4**	1.52	2.34	3.24	3.54
−2.2	−1.59	−2.64	−3.38	—	**5.8**	1.62	2.48	3.64	—
−1.8	−1.40	−2.44	−3.11	—	**6.2**	1.72	2.62	—	—
−1.4	−1.21	−2.22	−2.87	—	**6.6**	1.81	2.75	—	—
−1.0	−1.01	−1.96	−2.56	−3.72	**7.0**	1.90	2.87	—	—
−0.6	−0.80	−1.66	−2.20	−2.88	**7.4**	1.98	2.97	—	—
−0.2	−0.60	−1.31	−1.81	−2.27	**7.8**	2.07	3.08	—	—
0.2	−0.39	−0.94	−1.41	−1.85	**8.2**	2.15	3.22	—	—
0.6	−0.19	−0.57	−0.97	−1.38	**8.6**	2.23	3.36	—	—
1.0	0.00	−0.19	−0.51	−0.84	**9.0**	2.31	—	—	—
1.4	0.18	0.15	−0.06	−0.33	**9.4**	2.38	—	—	—
1.8	0.35	0.45	0.37	0.18	**9.8**	2.45	—	—	—

n	b_n	c_n	d_n	n	b_n	c_n	d_n
7	−2.356	1.245	0.4533	**29**	−6.074	1.934	0.1907
8	−2.696	1.333	0.4186	**30**	−6.150	1.949	0.1872
9	−2.968	1.400	0.3900				
10	−3.262	1.471	0.3600	**31**	−6.248	1.965	0.1840
				32	−6.324	1.976	0.1811
11	−3.485	1.515	0.3451	**33**	−6.402	1.988	0.1781
12	−3.731	1.571	0.3270	**34**	−6.480	2.000	0.1755
13	−3.936	1.613	0.3111	**35**	−6.559	2.012	0.1727
14	−4.155	1.655	0.2969				
15	−4.373	1.695	0.2842	**36**	−6.640	2.024	0.1702
				37	−6.721	2.037	0.1677
16	−4.567	1.724	0.2727	**38**	−6.803	2.049	0.1656
17	−4.713	1.739	0.2622	**39**	−6.887	2.062	0.1633
18	−4.885	1.770	0.2528	**40**	−6.961	2.075	0.1612
19	−5.018	1.786	0.2440				
20	−5.153	1.802	0.2359	**41**	−7.035	2.088	0.1591
				42	−7.111	2.101	0.1572
21	−5.291	1.818	0.2264	**43**	−7.188	2.114	0.1552
22	−5.413	1.835	0.2207	**44**	−7.266	2.128	0.1534
23	−5.508	1.848	0.2157	**45**	−7.345	2.141	0.1516
24	−5.605	1.862	0.2106				
25	−5.704	1.876	0.2063	**46**	−7.414	2.155	0.1499
				47	−7.484	2.169	0.1482
26	−5.803	1.890	0.2020	**48**	−7.555	2.183	0.1466
27	−5.905	1.905	0.1980	**49**	−7.615	2.198	0.1451
28	−5.988	1.919	0.1943	**50**	−7.677	2.212	0.1436

对 $3 \leqslant n \leqslant 6$，首先计算 $v = \ln[(T-d_n)/(1-T)]$，这里 d_n 由表的第二行给出，T 是 Shapiro-Wilk 检验统计量，然后用表中的数 v 和 n 去求出近似正态的 G.

来源：经 *Biometrika* 委托许可，由 Pearson 和 Hartley(1976)第二卷再版修改而来.

对 $7 \leqslant n \leqslant 50$，由上表中的 n，找出 b_n，c_n 和 d_n，然后计算

$$G = b_n + c_n \ln\{(T-d_n)/(1-T)\}$$

它是近似标准正态的.

表 A19　等样本容量 n 的两样本 Smirnov 检验统计量的分位数[a]

	单边检验: $p = 0.90$	0.95	0.975	0.99	0.995		单边检验: $p = 0.90$	0.95	0.975	0.99	0.995
	双边检验: $p = 0.80$	0.90	0.95	0.98	0.99		双边检验: $p = 0.80$	0.90	0.95	0.98	0.99
$n = 3$	2/3	2/3				$n = 22$	7/22	8/22	8/22	10/22	10/22
4	3/4	3/4	3/4			23	7/23	8/23	9/23	10/23	10/23
5	3/5	3/5	4/5	4/5	4/5	24	7/24	8/24	9/24	10/24	11/24
6	3/6	4/6	4/6	5/6	5/6	25	7/25	8/25	9/25	10/25	11/25
7	4/7	4/7	5/7	5/7	5/7	26	7/26	8/26	9/26	10/26	11/26
8	4/8	4/8	5/8	5/8	6/8	27	7/27	8/27	9/27	11/27	11/27
9	4/9	5/9	5/9	6/9	6/9	28	8/28	9/28	10/28	11/28	12/28
10	4/10	5/10	6/10	6/10	7/10	29	8/29	9/29	10/29	11/29	12/29
11	5/11	5/11	6/11	7/11	7/11	30	8/30	9/30	10/30	11/30	12/30
12	5/12	5/12	6/12	7/12	7/12	31	8/31	9/31	10/31	11/31	12/31
13	5/13	6/13	6/13	7/13	8/13	32	8/32	9/32	10/32	12/32	12/32
14	5/14	6/14	7/14	7/14	8/14	33	8/33	9/33	11/33	12/33	13/33
15	5/15	6/15	7/15	8/15	8/15	34	8/34	10/34	11/34	12/34	13/34
16	6/16	6/16	7/16	8/16	9/16	35	8/35	10/35	11/35	12/35	13/35
17	6/17	7/17	7/17	8/17	9/17	36	9/36	10/36	11/36	12/36	13/36
18	6/18	7/18	8/18	9/18	9/18	37	9/37	10/37	11/37	13/37	13/37
19	6/19	7/19	8/19	9/19	9/19	38	9/38	10/38	11/38	13/38	14/38
20	6/20	7/20	8/20	9/20	10/20	39	9/39	10/39	11/39	13/39	14/39
21	6/21	7/21	8/21	9/21	10/21	40	9/40	10/40	12/40	13/40	14/40
						$n > 40$ 的近似值	$\frac{1.52}{\sqrt{n}}$	$\frac{1.73}{\sqrt{n}}$	$\frac{1.92}{\sqrt{n}}$	$\frac{2.15}{\sqrt{n}}$	$\frac{2.30}{\sqrt{n}}$

来源：经 Institute of Mathematical Statistics 许可，由 Birnbaum 和 Hall（1960）修改而来.

a 表中的数是（6.3.1）式定义的 Smirnov 两样本双边检验统计量 T（（6.3.2）和（6.3.3）定义的两样本单边检验）有选择的 p 分位数 w_p. 如果统计量 T 超过表中给出的 $1-\alpha$ 分位数 $w_{1-\alpha}$，则以水平 α 拒绝 H_0. 检验统计量是离散型随机变量，所以精确的置信水平可能小于表中出现的 α.

表 A20　不等样本容量 n,m 的两样本 Smirnov 检验统计量的分位数[a]

单边检验：		$p = 0.90$	0.95	0.975	0.99	0.995
双边检验：		$p = 0.80$	0.90	0.95	0.99	0.99
$N_1 = 1$	$N_2 = 9$	17/18				
	10	9/10				
$N_1 = 2$	$N_2 = 3$	5/6				
	4	3/4				
	5	4/5	4/5			
	6	5/6	5/6			
	7	5/7	6/7			
	8	3/4	7/8	7/8		
	9	7/9	8/9	8/9		
	10	7/10	4/5	9/10		
$N_1 = 3$	$N_2 = 4$	3/4	3/4			
	5	2/3	4/5	4/5		
	6	2/3	2/3	5/6		
	7	2/3	5/7	6/7	6/7	
	8	5/8	3/4	3/4	7/8	
	9	2/3	2/3	7/9	8/9	8/9
	10	3/5	7/10	4/5	9/10	9/10
	12	7/12	2/3	3/4	5/6	11/12
$N_1 = 4$	$N_2 = 5$	3/5	3/4	4/5	4/5	
	6	7/12	2/3	3/4	5/6	5/6
	7	17/28	5/7	3/4	6/7	6/7
	8	5/8	5/8	3/4	7/8	7/8
	9	5/9	2/3	3/4	7/9	8/9
	10	11/20	13/20	7/10	4/5	4/5
	12	7/12	2/3	2/3	3/4	5/6
	16	9/16	5/8	11/16	3/4	13/16
$N_1 = 5$	$N_2 = 6$	3/5	2/3	2/3	5/6	5/6
	7	4/7	23/35	5/7	29/35	6/7
	8	11/20	5/8	27/40	4/5	4/5
	9	5/9	3/5	31/45	7/9	4/5
	10	1/2	3/5	7/10	7/10	4/5
	15	8/15	3/5	2/3	11/15	11/15
	20	1/2	11/20	3/5	7/10	3/4
$N_1 = 6$	$N_2 = 7$	23/42	4/7	29/42	5/7	5/6
	8	1/2	7/12	2/3	3/4	3/4
	9	1/2	5/9	2/3	13/18	7/9
	10	1/2	17/30	19/30	7/10	11/15
	12	1/2	7/12	7/12	2/3	3/4
	18	4/9	5/9	11/18	2/3	13/18
	24	11/24	1/2	7/12	5/8	2/3
$N_1 = 7$	$N_2 = 8$	27/56	33/56	5/8	41/56	3/4
	9	31/63	5/9	40/63	5/7	47/63
	10	33/70	39/70	43/70	7/10	5/7
	14	3/7	1/2	4/7	9/14	5/7
	28	3/7	13/28	15/28	17/28	9/14
$N_1 = 8$	$N_2 = 9$	4/9	13/24	5/8	2/3	3/4
	10	19/40	21/40	23/40	27/40	7/10
	12	11/24	1/2	7/12	5/8	2/3
	16	7/16	1/2	9/16	5/8	5/8
	32	13/32	7/16	1/2	9/16	19/32
$N_1 = 9$	$N_2 = 10$	7/15	1/2	26/45	2/3	31/45
	12	4/9	1/2	5/9	11/18	2/3
	15	19/45	22/45	8/15	3/5	29/45
	18	7/18	4/9	1/2	5/9	11/18
	36	13/36	5/12	17/36	19/36	5/9
$N_1 = 10$	$N_2 = 15$	2/5	7/15	1/2	17/30	19/30
	20	2/5	9/20	1/2	11/20	3/5
	40	7/20	2/5	9/20	1/2	—
$N_1 = 12$	$N_2 = 15$	23/60	9/20	1/2	11/20	7/12
	16	3/8	7/16	23/48	13/24	7/12
	18	13/36	5/12	17/36	19/36	5/9
	20	11/30	5/12	7/15	31/60	17/30
$N_1 = 15$	$N_2 = 20$	7/20	2/5	13/30	29/60	31/60
$N_1 = 16$	$N_2 = 20$	27/80	31/80	17/40	19/40	41/80
大样本近似		$1.07\sqrt{\frac{m+n}{mn}}$	$1.22\sqrt{\frac{m+n}{mn}}$	$1.36\sqrt{\frac{m+n}{mn}}$	$1.52\sqrt{\frac{m+n}{mn}}$	$1.63\sqrt{\frac{m+n}{mn}}$

来源：经 Institute of Mathematical Statistics 许可，由 Massey（1952）修改而来.

a 表中的数是由（6.3.1），（6.3.2）和（6.3.3）式定义的 Smirnov 两样本检验统计量 T 有选择的 p 分位数 w_p. 令样本容量较小的为 N_1，样本容量较大的为 N_2，如果统计量 T 超过表中给出的 $1-\alpha$ 分位数 $w_{1-\alpha}$，则以水平 α 拒绝 H_0. 对于表中没有包含的 n 和 m，使用表中最后一行给出的大样本近似值，或者对于 n，$m \leqslant 100$ 的情形，使用在 Harter 和 Owen（1970）中由 Kim 和 Jennrich 制作的精确表.

表 A21　*t* 分布[a]

自由度	$p = 0.6$	0.75	0.9	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.377	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

来源：经 *Biometrika* 委托许可，由 Pearson 和 Hartley(1976)第一卷再版修改而来.

a 表中的数是不同自由度下的 t 分布的 p 分位数 w_p 的值，对于 $p<0.5$，p 分位数 w_p 可以由式

$$w_p = -w_{1-p}$$

得到，注意，对所有的自由度，有 $w_{0.50}=0$.

表A22 自由度为k_1, k_2的F分布（0.75分位数）

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.28
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	1.27
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	132	1.31	1.30	1.29	1.28	1.26	1.25
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.25	1.24
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00

来源：经*Biometrika* Trustees许可，由Pearson和Hartley(1976)第一卷再版修改而来.

(0.90分位数) (续)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.68
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

(0.95分位数) (续)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.30	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

(0.975分位数) (续)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.80	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

(0.99分位数) (续)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

奇数号习题解答

第1章

1.1节 1. 10,000; 3. 24; 5. 220; 7. 105; 9. 14; 11. 232/729.

1.2节 1. HHH,HHT,HTH,THH,TTH,THT,HTT,TTT; 3. 0.85; 5. $9/2^8$; 7. 0.007125; 9. 0.06; 11. 1/7; 13. 1/4; 15. 4/15; 17. (a)6/1000，若3个数字不同；3/1000，若一个数字出现两次；或1/1000，若一个数字出现3次. (b)否.

1.3节 1. (a)1/729, (b)64/729, (c)0, (d)496/729, (e)0, (f)1; 3. (a)1/128, (b)4/128, (c)3/128, (d)1/128, (e)0, (f)1/128, (g)12/128, (h)1; 5. (a)A, B, C, D, E, F, (b)$P(A)=1/6, P(B)=1/6, \cdots, P(F)=1/6$, (c) $X(A)=1, X(B)=2$, 等.

1.4节 1. (a)7/6, (b)41/36, (c)29/6, (d)1, (e)0.5, (f)1; 3. (a)1/2, (b)1/2, (c)1/4, (d)1, (e)0, (f)1/2, (g) 1/2, (h) 是; 5. 2211; 7. (a)7/2, (b)35/12, (c) 56/3; 9. $\mu=11$, $\sigma^2=44/3$, 极差 = 16; 11. (a)$P(X=0)=2/3, P(X=1)=1/3$, (b)0, (c) −1/9, (d) −1/2; 13. $\mu=13.5$, $\sigma^2=11.25$.

1.5节 1. (a)0.5, (b)0.975, (c)0.159, (d)0.682, (e)0.5, (f)0.6745; 3. 0.001; 5. 207;7. (a)9.488, (b)15.51, (c)233.7; 9. 0.266; 11. <0.0005.

第2章

2.1节 1. (a)美国的所有高中，(b)在华盛顿地区的所有高中，(c)75,287,520，(d)1/75,287,520; 3. (a)有序的，(b)所赢得奖励点的和；5. 所有名义的；7. 否.

2.2节 1. (b) 4300美元，(c)8600美元，(d)97,296 000美元，(e) 9863.87美元；3. 12.6, 12.1,0.5; 5. “样本值$\leqslant c$”的个数$/n$,1/5; 7. 1005美元到1262美元；9. 0.05到0.55; 11. $\hat{P}(x)=1$, x从0到187; 6/7, x从187到196; 9/14,x从196到206; 3/7,x从206到210; 3/14,x从210到273; 0,x大于273.

2.3节 1. (a)H_0: 新方法不比现有的方法好，H_1: 新方法比现有的方法好，(b)当它不好，决定新方法好的概率，(c)当它是好的，决定新方法好的概率；3. (a)肥料B不如肥料A好，(b)我的对手是骗子，(c)太阳黑子的出现不会影响经济周期；5. 功效 = 0.33696; 7. 0.10,0.91.

2.4节 1. (a)1/32, (b)p^5, $p>0.5$, (d)是；5. (a)0.57, (b)1.75.

第3章

3.1节 1. 不，$T=0$, $p=0.036$; 3. (a) 可能是，$p=0.055$, (b) (0.15, 0.27);

5. (0.057,0.437); 7. (a) (0.087,0.491), (b) (0.060,0.440); 9. $p=0.991$; 11. (a) (0.544,0.896), (b) (0.591,0.909).

3.2 节 1. $T_1=6, T_2=4, p=0.1154$; 3. $T_1=4, p=0.2375$; 5. 不, $T_1=90, T_2=90, p=0.022$; 7. $T_1=10, p<0.00001$.

3.3 节 1. (a)77, (b)76.35, 所以用 77; 3. (a)15, (b)14.198, 所以用 15; 5. 26; 7. (a)94.6%,95.6%, (b)91.2%,92.4%; 9. 210.

3.4 节 1. $T=1, p=0.1094$; 3. $T=23, p<0.0002$; 5. $T=60, p=0.046$.

3.5 节 1. $T_1=0.5455, p>0.25$; 3. $T=7, p=0.0078$; 5. $T=4, p=0.006$; 7. $T=6$, $p=0.0312$; 9. $T_2=3$, $n=17, p=0.0128$.

第 4 章

4.1 节 1. $T_1=0.800, p=0.424$; 3. $T_1=0.935, p=0.35$; 5. $T_2=1, p=0.005$; 7. (a)$T_4=0.8482$, 有常数系数, $p=0.198$, (b)$T_5=1.0599, p=0.145$, (c)T_5.

4.2 节 1. 2×3 具有列总和相等 ($R_1=8, R_2=10$), $T=6.3, p<0.05$; 3. $T=6.81, 0.05<p<0.10$, 对固定的边缘总和; 5. 组合 C 和 B, $T=2.496, 0.10<p<0.25$; 7. $T=6.605, p>0.25$.

4.3 节 1. 精确地 $T=1.10$, 近似地 $T=1.20, p>0.25$; 3. $T=34.875, p<0.001$; 5. $T=5.067, 0.05<p<0.10$.

4.4 节 1. (a)6.34, (b)0.178, (c)0.0317, (d)0.244, (e)0.0634; (f) 0.1780; 3. (a)$R_5=0.149, p=0.117$, (b)0.3636, (c) 0.

4.5 节 1. 答案依赖于区间的选择; 3. 从 4.2 节, $T=19.03, p>0.25$; 5. $T=4.247, 0.10<p<0.25$.

4.6 节 1. (a) $T=4, 0.025<p<0.05$, (b)$T=4, 0.025<p<0.05$, (c) $T=2.67, 0.10<p<0.25$; 3. $T=10.4, 0.001<p<0.005$.

第 5 章

5.1 节 1. $T=117, T'=45$, $p<0.001$; 3. (0,14); 5. $T=24, T'=12, 0.05<p<0.10$, 精确地 $p=4/70=0.057$.

5.2 节 1. $T=8.40$, $p<0.01$ 从表 A8, 在 $\alpha=0.05$ 水平下, A 与 B, A 与 C 有差异; 3. 在 $\alpha=0.05$ 水平下, 最小耕作和梯田, 最小耕作和其他, 等高地型与梯田, 等高地型与其他有差异; 5. 在 $\alpha=0.05$ 水平下, 除了 B 和 E 外, 所有的配对有差异.

5.3 节 1. $T=1107.5, 0.05<p<0.10$; 3. $T_2=5.15, 0.05<p<0.10$.

5.4 节 1. (a) −0.603, (b) −0.511, (c)$0.05<p<0.10$, (d)$0.02<p<0.05$; 3. (a) $\rho=-0.9025, p<0.001$, (b) $T=-57, p<0.005$; 5. (a)0.8721, $0.025<p<0.05$, (b) 0.7778, $0.025<p<0.05$; 7. $N_c=98, N_d=27, \tau=0.5680, p=0.014$.

5.5 节 1. (b)$a=8.69, b=13.6$, (d)$\rho=-0.6606, 0.02<p<0.05$, (e)$(S^{(12)}, S^{(34)})=(10.7, 17.4)$; 3. $\rho=-0.8373, p<0.001$.

5.6 节 1. (a) 在中间可能是线性的, 它当然像单调的, (b)46.14, (c)53.51, (d)$R(y)=0.4333+0.9212R(x)$, 依次连接 $(X,Y)=(0.50,0),(0.85,0),(1,2.76),(1.5,$

11.97),(1.94,20),(2,21.58),(2.5,33.86),(2.75,40),(3,46.14),(3.5,58.42),(3.56,60),(4,76.06),(4.11,80),(4.5,89.66),(4.92,100),(5.00,100)；3. $R(y)=3.9375+0.5625R(x)$，依次连接 (X,Y) =(4.0,0),(4.36,0),(4.4,0.01),(4.7,0.07),(5.1,0.14),(9.3,0.21),(10.7,0.27),(11.2,0.34),(13.7,0.35),(13.9,0.47),(15.0,0.54),(15.8,0.60),(17.0,0.67),(18.1,0.74),(19.7,0.80),(19.9,0.87),(20.7,0.93),(21.6,1.0).

5.7 节 1. $T=2.988, p=0.003$；3. $T^+=39, p\approx0.50$；5. 5.5 到 7.45,6.5；7. (a) −18 到 23，(b)61/64 =95.3%.

5.8 节 1. (a)$T_3=4.43$, $0.01<p<0.025$，春季与其他三个季节有差异，(b)$T_2=2.946$, $0.05<p<0.10$，(c)在第一次测试中，较大的医院有更多的权重；3. $T_2=8.328$, $0.05<p<0.10$；5. $T_5=4.736, p<0.0001$.

5.9 节 1. 轮胎类型 3 和 6 比其他的要好，轮胎类型 5 和 7 比轮胎类型 2 要好；3. 队 5 比队 1，2 和 4 要好，队 3 比队 1 和 2 要好，队 4 比队 2 要好.

5.10 节 1. $T_1=7.97, p=0.020$，与 K-W 有相同的差异；3. $T_2=-2.938, p=0.004$；5. $T_3=1.7140, p=0.043$；7. $\rho=-0.6925$, ρ 的绝对值较大，$p=0.038$，稍有些小.

5.11 节 1. $T_1=5, p=6/252=0.0238$；3. $T_2=7, p=8/256=0.03125$.

第 6 章

6.1 节 1. $S_n(x)\pm0.509$；3. $T=0.425, p>0.20$；5. $T=0.492, p<0.01$.

6.2 节 1. $T_1=0.103, p>0.20$；3. $T_3=0.9569, p>0.50$；5. $T_2=0.2155, p\approx0.20$；7. $Z=-1.9083, p=0.028$.

6.3 节 1. $T_1^+\doteqdot0.8, p=0.40$；3. $T_2=0.556$, $0.01<p<0.05$；5. $T_1=\frac{6}{10}, p=0.052$.

索　　引

索引中的页码为英文原书页码,与书中页边标注的页码一致.

A

A. R. E. (asymptotic relative efficiency)(渐近相对效率),112
- of Cox and Stuart test(Cox 和 Stuart 检验),175,323
- of Daniel test(Daniel 检验),322
- of Durbin test(Durbin 检验),394
- of Friedman test(Friedman 检验),379
- of Kendall's tau(Kendall τ),327
- of Kruskal-Wallis test(Kruskal-Wallis 检验),297
- of Mann-Whitney test(Mann-Whitney 检验),284
- of median test(中位数检验),285,297
- of paired t test(配对 t 检验),364
- of Quade test(Quade 检验),380
- of quantile test vs. one-sample t test(分位数检验对一样本 t 检验),148
- of rank test for slope(斜率的秩检验),342
- of sign test(符号检验),363
- of sign test vs. t test(符号检验与 t 检验),164,175
- of sign test vs. Wilcoxon signed ranks test(符号检验与 Wilcoxon 符号秩检验),164,175
- of Spearman's rho(Spearman ρ),327
- of squared ranks test(平方秩检验),309
- of two-sample t test(两样本 t 检验),284
- of Wilcoxon signed ranks test(Wilcoxon 符号秩检验),363

acceptance region(接受域),98
aligned-rank methods(秩排列方法),384,385
alternative hypothesis(备择假设),95
alternatives,ordered(备择的,次序的),297
analysis of covariance(方差分析),297
analysis of variance,one-way(方差分析,一种方式),222,297
approximate confidence interval for μ(μ 的近似置信区间),85
approximation formulas for tolerance limits(容忍限逼近公式),151,155
approximation,normal(正态,逼近):
- to binomial distribution(二项分布),58
- to sum of ranks(秩和),58

approximations to chi-squared distribution(χ^2 分布近似),62
asymptotic relative efficiency,(参见 A. R. E.)(渐近相对效率),112
asymptotically distribution-free methods(渐近分布自由方法),117

B

biased estimators for σ(σ 的有偏估计量),85
biased test(有偏检验),108
binomial coefficient(二项系数),9,11
binomial distribution(二项分布),28
- mean and variance in(均值和方差),49
- normal approximation to(正态逼近),58
- tables of the(表格),513-524
- tests based on the(基于……的检验),123

binomial expansion(二项展开),11
binomial test(二项检验),104,124
- power of(功效),127

bioassay(生物鉴定),119

bivariate random variable(二维随机变量),72
block design, incomplete(不完全的区组设计),387
randomized complete(完全随机化),251,368
blocks, multiple comparisons with complete(完全区组多重比较),371,375
bootstrap(自助法),349
bootstrap method of estimation(估计的 bootstrap 方法),86

C

censored data(删失数据),297
censored samples(删失样本),155,285
central limit theorem(中心极限定理),57,85
centroid(重心),36
chi-squared approximation to Kruskal-Wallis test (χ^2 近似 Kruskal-Wallis 检验),295
chi-squared distribution function(χ^2 分布函数的),54,59
approximations to(逼近到),62
tables(表格),512
chi-squared goodness-of-fit test(χ^2 拟合优度检验),239,240,429,430,442,443
chi-squared random variables, sum of(χ^2 随机变量的和),62
chi-squared test(χ^2 检验):
for differences in probabilities(概率差异),180,199
with fixed marginal totals(固定边缘总和),209
for independence(独立性),204
circular distributions(圆周分布),285,364
cluster analysis(聚类分析),419
Cochran test(Cochran 检验),250
Cochran's criteria for small expected values(对小期望值的 Cochran 准则),202
coefficient, binomial(系数,二项),9,11
confidence(置信),83
multinomial(多项式),9,12
coefficient of concordance, Kendall's (Kendall 一致性系数),328,380
comparisons, multiple(多重对比):
with complete blocks(完全区组),371,375
incomplete blocks(不完全区组),390
with independent samples(独立样本),290,297,398
in test for variances(方差检验),304
complete block design, randomized(随机化完全区组设计),368
completely randomized design(完全随机化设计),222
composite hypothesis(复合假设),97
computer simulation to find null distribution(计算机模拟求零假设分布),446,447
concordance between blocked rankings(区组秩间的一致性),385
concordance, Kendall's coefficient of (Kendall 系数一致性),328,382
concordant pairs(不和谐配对),319
conditional probability(条件概率),17,23,24
conditional probability function(条件概率函数),29
confidence band for a distribution function(分布函数置信界),438
confidence coefficient(置信系数),83,129,143
confidence interval(系数,区间),83,114,129
for the difference between two means(两均值差异),281
for a mean, parametric(均值,参数),149
for the median difference(中位差异),360
for μ, approximate(对于 μ,逼近),85
for a probability or population proportion(对于概率或总体比例),130
exact tables for(精确表格),525-536
for a quantile(分位数),135,143
one-sided(单边),153
for a slope(斜率),335

conservative test(保守检验),113
consistent(相合的),117
consistent sequence of tests(检验的相合序列),106,108,160
consistent, sign test(相合,符号检验),163
contingency coefficient(列联系数):
Cramér's(Cramér),229
Pearson's(Pearson),231
Pearson's mean square(Pearson 均值平方),231
contingency table(列联表),166,179,199,292
fourfold(四重的),180
multi-dimensional(多维的),215
$r \times c$($r \times c$ 维),199
three-way(三种方式的),214
two-way(两种方式的),214
continuity correction(连续修正),126,127,135,138,159,190,192,194,195
in Kendall's tau(Kendall's τ),322
in Mann-Whitney test(Mann-Whitney 检验),274,275
in Wilcoxon signed ranks test(Wilcoxon 符号秩检验),359
continuous distribution function(连续分布函数),53
continuous random variable(连续型随机变量),52,53
control, sign test for comparing several treatments with a(带控制的几种处理比较的符号检验),175
convenience sample(方便样本),69
correction for continuity(连续修正),126,127,135,138,159,190,192,194,195
correction, Sheppard's(Sheppard 修正),248
correlation(相关性):
quick test for(快速检验),196
rank(秩),312
sign test for(符号检验),172
correlation coefficient(相关系数):
Kendall's partial(Kendall 偏),327
Kendall's rank(Kendall 秩),318,319,325,326
Pearson's product moment(Pearson 乘积矩),313,318
Spearman's rank(Spearman 秩),314,325,326
correlation coefficient between two random variables(两随机变量的相关系数),43
correlation test(相关性检验):
Kendall's rank(Kendall 秩),175,321
Spearman's rank(Spearman 秩),175,316
counting rules(计数法则),5
covariance(协方差),39
analysis of(分析),297
of two random variables(两随机变量),41,42,46
of two ranks(两秩),45
Cox and Stuart test for trend(Cox 和 Stuart 趋势检验),169,170
A. R. E. of(A. R. E. 的),175
Cramér's coefficient(Cramér 系数),230,234
Cramér's contingency coefficient(Cramér 列联系数),229
Cramér-von Mises goodness-of-fit test(Cramér'-von Mises 拟合优度检验),441
Cramér-von Mises two-sample test(Cramér'-von Mises 两样本检验),463
tables for(表格),464
critical region(临界区域),97,98,101
size of(大小),100
curves, survival(生存曲线),119

D

Daniel's test for trend(Daniel 趋势检验),323
decile,(十分位数(的)) 33,34
decision rule(决策法则),98
degrees of freedom(自由度),59
dependence, measure of(相依性度量),227

design(设计):
completely randomized(完全随机化),222
experimental(经验),419
incomplete block(不完全区组),387
randomized complete block(随机化完全区组),368
deviates,random normal(正态随机偏离),404
difference between two means,confidence interval for the(两均值差异的置信区间),281
difference, confidence interval for the median(中位数差异置信区间),360
discordant pairs(不和谐配对),319
discrete distribution function(离散分布函数),52
discrete random variable(离散型随机变量),52
discrete uniform distribution(离散均匀分布),28,437
discriminant analysis(判别分析),119,419
dispersion,sign test for trends in(散布趋势的符号检验),175
distribution(分布):
binomial(二项),27
discrete uniform(离散均匀),28
exponential(指数),447
hypergeometric(超几何分布),30
lognormal(对数正态分布),453
null(零),99
uniform(均匀),433
distribution-free(分布自由),114
distribution-free methods, asymptotically(渐近的分布自由方法),117
distribution function(分布函数),26
chi-squared(χ^2),54,59
confidence band for(置信界),438
continuous(连续),53
discrete(离散),52
empirical(经验),79,428
joint(联合),29
normal(正态),54,55
of order statistics(次序统计量),146,147,153
sample(样本),79,80
distributions with heavy tails(重尾分布),116,148,164
distributions with light tails(轻尾分布),116,164
dose-response curves(剂量响应曲线),349
Durbin test(Durbin 检验),387,388
efficiency of(效率),394

E

efficiency(效率),106
asymptotic(渐近的),112
of the Durbin test(Durbin 检验),394
of the Friedman test(Friedman 检验),379
of the paired t-test(配对 t 检验的),364
relative(相关的),110,111,112
of the sign test(符号检验),364
of the Smirnov test(Smirnov 检验),465
of the Wilcoxon test(Wilcoxon 检验),364
empirical distribution function(经验分布函数),79,428
empirical survival function(经验生存函数),89
empty set(空集),14
error(误差):
standard(标准),85,88
type Ⅰ(Ⅰ类),98
type Ⅱ(Ⅱ类),98,99
estimate(估计):
interval(区间),83
point(点),83
of the standard deviation(标准差),443
estimation(估计),79,88
of parameters in chi-squared goodness-of-fit test(参数 χ^2 拟合检验),243,245,249
estimator(估计量),79,81
of population mean(样本均值),115

of population standard deviation(样本标准差),115
unbiased(无偏),74
for $\mu(\mu)$,84
for $\sigma^2(\sigma^2)$,85
event(事件),7,14
probability of(概率),14
sure(必然),14
events,independent(事件,独立),18,19
joint(联合),17
mutually exclusive(互不相容),19
exact test,Fisher's(精确检验,Fisher),188,213
exclusive,mutually(互不相容),14
expected normal scores(期望正态得分),404
expected value(期望(值)),35,39
expected values,small(小期望(值)):
in contingency tables(列联表),201,220
in goodness-of-fit test(拟合优度检验),241,249
experiment(试验),6,69
experimental design(试验设计),419
experiments,independent(独立试验),15,19,20
exponential distribution(指数分布),447
Lilliefors test for the(Lilliefors 检验),448
extension of the median test(中位数检验的扩展),224

F

*F*distribution(*F* 分布):
in Friedman test(Friedman 检验),370
in incomplete block analysis(不完全区组分析),389
in Quade test(Quade 检验),374
table of the(表格),562-571
F statistic(*F* 统计量),227,258,418
computed on scores(得分计算),312
F test(*F* 检验),297,300
for equal variances(等方差),308,309
for randomized complete blocks(随机完全区组),379
factorial notation(阶乘记号),8
families of distributions,goodness-of-fit tests for(分布族的拟合优度检验),442
Fisher's(Fisher):
exact test(精确检验),188,213
least significant difference(最小显著差异),296
LSD procedure on ranks(有关秩的 LSD 方法),379
method of randomization(随机化方法),407
four-fold contingency table(四重列联表),180,233
freedom,degrees of(自由度),59
Friedman test(Friedman 检验),367,369
efficiency of(效率),379
extension of(推广),383
function(函数):
distribution(分布),26
power(功效),163
probability,of a random variable(随机变量的概率),25
probability,on a sample space(样本空间的概率),15
random(随机),80
step(阶梯),52
survival(生存),89

G

gamma coefficient(gamma 系数),320
goodness-of-fit test(拟合优度检验):
chi-squared(χ^2),239,240
Cramér-von Mises(Cramér-von Mises),441
Kolmogorov(Kolmogorov),428,430,435
goodness-of-fit tests for families of distributions(分布族拟合优度检验),442
grand median(全中位数),218

H

heavy tails,distributions with(重尾分布),116,

148,164
Hodges-Lehman estimate of shift(Hodges-Lehman 漂移估计),282,361
hypergeometric distribution(超几何分布),30,188,191
mean of(均值),188,191
standard deviation of(标准差),188,191
hypothesis(假设):
alternative(备择的),95
composite(复合的),97
null(零),95
simple(简单),97
testing(检验),95
tests,properties of(检验,性质),106

I

incomplete block design(不完全区组设计),368,387
incomplete blocks,multiple comparisons(不完全区组的多重比较),390
independence,the chi-squared test for(独立,χ^2 检验),204
independent(独立):
events(事件),18,19
experiments(试验),15,19,20
random variables(随机变量),31,46,72
samples,multiple comparisons with(样本,多重比较),290,297,398
samples,randomization test for two(样本,随机化检验),409
inference,statistical(统计推断),68
interaction(交互):
rank transformation test for(秩变换检验),419
test for(检验),384
intercept(截距),333
interquartile range(四分位数极差),37
interval,confidence(置信区间),83
interval estimate(区间估计),83,129
interval scale of measurement(测量的区间尺度),74

J

joint distribution function(联合分布函数),28,29
joint event(联合事件),17
joint probability function(联合概率函数),28
Jonckheere-Terpstra test for ordered alternatives(Jonckheere-Terpstra 顺序备择检验),325

K

Kaplan-Meier estimator(Kaplan-Meier 估计量),89
Kendall's(Kendall):
coefficient of concordance(一致性系数),328
partial correlation coefficient(偏相关系数),327
rank correlation test(秩相关检验),175,321
tau (τ),318,319,325,326,335
exact tables(精确表),545-546
tau(τ),A. R. E. of,327
tau for ordered alternatives(顺序备择 τ),381
Klotz test(Klotz 检验),401
Kolmogorov goodness-of-fit test(Kolmogorov 拟合优度检验),428,430
exact tables(精确表),549
Kolmogorov goodness-of-fit test for discrete distributions(离散分布的 Kolmogorov 拟合优度检验),435
Kolmogorov-Smirnov tests(Kolmogorov-Smirnov 检验),428
Kruskal-Wallis test(Kruskal-Wallis 检验),288
exact tables for(精确表),541

L

least significant difference,Fisher's(最小显著

差异,Fisher),296
least squares estimates(最小二乘估计),334
least squares method(最小二乘方法),333
Let's make a deal(让我们和……妥协),66
level of significance(显著水平),99
life testing(寿命检验),148
light tails, distributions with(轻尾分布),116,164
likelihood ratio statistic(似然比统计量),258
likelihood ratio test(似然比检验),259
Liliefors test for the exponential distribution(指数分布的 Liliefors 检验),448
table(表格),551
Liliefors test for normality(Liliefors 正态性检验),443
tables(表格),550
limits,tolerance(容忍限),150
linear regression(线性回归),333
location estimates, robust(位置估计,稳健),362
location,measure of(位置,度量),36
loglinear models(对数线性模型),215,259
lognormal distribution(对数正态分布),453
longitudinal studies(纵向研究),119
lottery game,Texas Lotto(Texas 彩票游戏),66
lower-tailed test(左边检验),98

M

Mann-Whitney test(Mann-Whitney 检验),103,203,271
tables(表格),538-540
Mantel-Haenszel test(Mantel-Haenszel 检验),192
marginal totals, chi-squared test with fixed(固定边缘和的χ^2 检验),209
matched pairs(配对),350
randomization test for(随机化检验),412
McNemar test(McNemar 检验),166,180,252,255,256
compared with paired t test(与配对 t 检验比较),178
mean(均值),36,51
of hypergeometric distribution(超几何分布),188,191
population,estimator of(总体估计量),115
in rank test using scores(得分的秩检验),306
sample(样本),81,83
of sum of random variables(随机变量和),39
of sum of ranks(秩和),41,49
and variance in binomial distribution(二项分布的方差),49
means(均值):
confidence interval for the difference between two(两差异的置信区间),281
sign test for equal(对相等的符号检验),160
measurement scale(度量尺度),73
interval(区间),74
nominal(名义),73
ordinal(有序的),74
ratio(比率),75
measures of dependence(相依度量),227
median(中位数),33,34
difference, confidence interval for(差异,置信区间),360
grand(总的),218
sample(样本),82
test(检验),218,352,355
comparison with Kruskal-Wallis test(与 Kruskal-Wallis 检验的比较),291
an extension of(一个推广),224
medians,sign test for equal(对相等中位数的符号检验),160
meta-analysis(meta-分析),452
minimum chi-squared method(最小χ^2 距离方法),243,245

Minitab(Minitab),91,107,127,130,139,144,161,182,201,205,210,220,241,276,282,290,318,322,328,336,355,361,371,382,390,444,451
model(模型),6
models,loglinear(对数线性模型),215,259
monotonic regression(单调回归),344
Mood test for variances(Mood 方差检验),309,312
multi-dimensional contingency table(多维列联表),215
multinomial(多项式的):
  coefficient(系数),9,12
  distribution(分布),203,207,249
  proportions,simultaneous confidence intervals for(比例,联合置信区间),133
multiple comparisons(多重比较):
  complete blocks design(完全区组设计),371,375
  incomplete blocks design(不完全区组设计),390
  independent samples(独立样本),290,297,398
  in one-way layout(以一种方式设计),220,222,252
  variance(方差),304
multiple regression(多元回归),419
multivariate data,randomization test for(多元数据的随机化检验),416
multivariate observations(多元观察),385
multivariate random variable(多元随机变量),71,72
  confidence region for(置信区间),362,364
mutually exclusive(互不相容),14
  events(事件),19

N

nominal scale data(名义尺度数据),117,118
nominal scale of measurement(测量的名义尺度),73
nonparametric methods(非参数方法),116
nonparametric statistics(非参数统计),2,114
  definition(定义),118
normal approximation(正态逼近):
  to binomial distribution(二项分布),58
  to chi-squared distribution(χ^2 分布),62
  to hypergeometric(超几何),188,194
  in Mann-Whitney test(Mann-Whitney 检验),237,281
  in squared ranks test(秩平方检验),301,302
  to sum of ranks(秩和),58
  in Wilcoxon signed ranks test(Wilcoxon 秩和检验),353,359
normal deviates,random(随机正态偏差),404
normal distribution function(正态分布函数),54,55
  standard(标准正态分布函数),55
    tables of(标准正态分布函数表),508-511
normal scores(标准得分),396
  expected(期望的),404
  in matched pairs test(配对检验),400
  in one-way layout(以一种方式设计),397
  in test for correlation(相关检验),403
  in test for variances(方差检验),401
  in two-way layout(以两种方式设计),403
normality(正态):
  Lilliefors test for(Lilliefors 正态检验),443
  Shapiro-Wilk test for(Shapiro-Wilk 检验),450
normalized sample(标准化样本),443
null distribution(零分布),99
null hypothesis(零假设),96

O

one-sample case(一样本情形),350
one-sample *t*-test(一样本 *t* 检验),363,418

one-tailed test(单边检验),98
one-to-one correspondence(一一对应),52
one-way analysis of variance(一种方式的方差分析),222,297
one-way layout(一种方式的设计),227
order statistic of rank k(秩 k 的次序统计量),77,82
order statistics(次序统计量),143
distribution function of(分布函数),146,147,153
ordered alternatives(有序备择),297,385
Jonckheere-Terpstra test for (Jonckheere-Terpstra 检验),325
Page test for(Page 检验),380
ordered categories,analysis of contingency table with(有序分类的列联表分析),292
ordered observation(次序观察),77
ordered random sample(次序随机样本),77
ordinal data(有序数据),117,118,271,272
ordinal scale of measurement(测量的顺序尺度),74
outcomes(结果),6
outliers(离群值),117,284,297

P

p-value(p-值),101
Page test for ordered alternatives(有序备择的 Page 检验),380
paired t test(配对 t 检验),363
efficiency of(效率),364
McNemar test compared with (McNemar 检验的比较),178
parallelism of two regression lines(两回归直线的平行),364
parameter estimation(参数估计),88
parametric confidence interval for mean(均值的参数置信区间),149
parametric methods(参数方法),115
parametric statistics(参数统计),2,114
partial correlation coefficient(偏相关系数):
Kendall's(Kendall),327
Spearman's(Spearman),328
PASS,107
Pearson product moment correlation coefficient (Pearson 乘积矩相关系数),234,239
Pearson's(Pearson)
contingency coefficient(列联系数),231
mean-square contingency coefficient(均方列联系数),231
product moment correlation coefficient(乘积矩相关系数),313,318
percentile(百分位点),33,34
phi coefficient(phi 系数),234,239
Pitman's efficiency(Pitman 有效性),112
point estimate(点估计),83
point in the sample space(样本空间中的点),13
population(总体),68,69
sampled(抽样),69,70
target(目标),69,70
power(功效),3,100,106,116
of the binomial test(二项检验),127
function(函数),106,163
probabilities,chi-squared test for differences in (概率差异的χ^2 检验),180,199
probability(概率),5,13
conditional(条件的),17,23
confidence interval for(置信区间),130
of the event(事件),14
function(函数),15
conditional(条件的),29
joint(联合),28
of the point(点的),14
sample(样本),69
properties of random variables(随机变量的性质),33
proportion, confidence interval for population (比例,总体的置信区间),130

Q

Quade test(Quade 检验),367,373
  efficiency of(效率),380
  power of(功效),380
quantile(分位数),27,33,34
  confidence interval for(置信区间),135,143
  population(总体),136
  sample(样本),81
  test(检验),135,136,222
    A. R. E. vs. one-sample t test(A. R. E. 与一样本 t 检验),148
quartile(四分位数),33,34

R

random function(随机函数),80
random normal deviates(随机正态偏差),404
random sample(随机样本),69,70,71
random variable(随机变量),22,23,76
  bivariate(二维),72
  continuous(连续),52,53
  discrete(离散),52
  distribution function of(分布函数),26
  multivariate(多元),71,72
  probability function of(概率函数),25
random variables(随机变量):
  correlation coefficient between two(两随机变量的相关系数),43
  covariance of two(两随机变量的协方差),41,42,46
  independent(独立),31,46,72
  properties of(性质),33
randomization, Fisher's method of(随机化的 Fisher 方法),407
randomization test for two independent samples(两独立样本的随机化检验),409
randomized complete block design(随机化完全区组设计),251,368
randomness, test for(随机检验),242
range(极差),37
  interquartile(四分位数间的),37
rank correlation(秩相关),312
  Kendall's test for(Kendall 检验),175,321
  Spearman's test for(Spearman 检验),175,316
rank of an order statistic(次序统计量的秩),77
rank transformation(秩变换),417
ranks(秩):
  covariance of two(两随机变量的协方差),45
  mean of sum of(和的均值),41,49
  normal approximation to sum of(和的正态近似),58
  variance of sum of(和的方差),48,49
ratio scale of measurement(测量的比率尺度),75
region(域):
  acceptance(接受),98
  critical(临界),97,98,101
  rejection(拒绝),98
regression(回归),328,332
  equation(方程),332
  linear(线性),333
  monotonic(单调),344
  multiple(多元),419
  parallelism of two lines(两线平行),364
rejection region(拒绝域),98
relative efficiency(相对效率),110,111,112
  asymptotic(渐近),112
Resampling Stats,88
research hypothesis(假设研究),95
rho, Spearman's(ρ, Spearman),314,325,326,335
  relationship with Friedman's test(与 Friedman 检验的关系),382
robust(稳健),419,420
  location estimates(局部估计),362

methods(方法),115,119
runs tests(游动检验),3

S

S -Plus,88,91,127,130,168,182,189,193,201,205,210,241,276,290,318,322,355,371,432,444,449
sample(样本),68,69
censored(删失),155,285
convenience(方便),69
distribution function(分布函数),79,80
mean(均值),81,83
mean, unbiased for μ(均值,对 μ 无偏),84
median(中位数),82
normalized(正则化),443
probability(概率),69
quantile(分位数),81
sequential(序贯),362
space(空间),13
point in the(样本空间中的点),13
standard deviation(标准差),83
variance(方差),81,83
unbiased for σ^2(对 σ^2 无偏),85
sampled population(抽样总体),69,70
SAS,168,182,189,193,201,205,210,230,259,276,290,322,325,355,371,390,451
scale,measure of(度量,刻度(尺度)),37
scale,tests for(刻度,检验),309,310
scales,measurement(刻度,测量),73
scores(得分),306
expected normal(期望正态),404
F statistic computed on(计算 F 统计量),312
mean in rank test using(均值的秩检验),306
normal(正态),396
variance in rank test using(方差的秩检验),307
sequential sampling(序贯抽样),362
sequential testing(序贯检验),285
set,empty(空集),14
Shapiro-Wilk test for normality(Shapiro-Wilk 正态检验),450
tables(表),552-557
Sheppard's correction(Sheppard 相关),248
Siegel-Tukey test(Siegel-Tukey 检验),312
sign test(符号检验),157
consistent(相合),163
for correlation(相关性),172
efficiency of(效率),364
for equal means(等均值),160
for equal medians(等中位数),160
extension to k samples of(推广到 k 样本),367
unbiased(无偏),163
variations of(方差),166,175
vs. t test, A. R. E. of(与 t 检验的 A. R. E.),164,175
vs. Wilcoxon signed ranks test, A. R. E of(与 Wilcoxon 符号秩检验的 A. R. E.),164,175
signed ranks test, Wilcoxon(Wilcoxon 符号秩检验),352
significance, level of(显著水平),99
simple hypothesis(简单假设),97
simulation, computer, to find null distribution(求零分布的计算机模拟),446,447
simultaneous confidence intervals(联合置信区间),133
size of the critical region(临界域的大小),100
slope, A. R. E. of rank test for(斜率的 A. R. E. 秩检验),342
slope in linear regression(线性回归的斜率),333
confidence interval for(置信区间),335
testing the(检验),335
Smirnov test(Smirnov 检验),456

efficiency of(效率),465
exact tables(精确表),558-560
Smirnov-type tests for several samples(多样本 Smirnov 型检验),462
Spearman's footrule(Spearman 脚规则),331
Spearman's rank correlation test(Spearman 秩相关检验),175,316
A. R. E. of(A. R. E.),327
exact tables(精确表),544
Spearman's rho (Spearman ρ), 314, 325, 326,335
for ordered alternatives(顺序备择),380
relationship with Friedman's test(与 Friedman 检验的关系),382
split plots(裂区),385
SPSS,v,382,390
squared ranks test for variances(方差的平方秩检验),300
exact tables for(精确表),542-543
standard deviation(标准差),37,38
estimate of(估计),443
of hypergeometric distribution(超几何分布),188,191
population, estimator of (总体估计量),115
sample(样本),83
standard error(标准差),85,88
standard normal distribution (标准正态分布),55
STATA,v,88
statistic(统计量),75,76
order(次序),77,82
test(检验),35,96,97
statistical inference(统计推断),68
statistics(统计学),68
StatMost,259
StatXact, 104, 127, 130, 144, 161, 168, 182, 189,201,205,210,220,230,241,252,276, 282,290,303,318,322,325,355,361,371, 375,380,382,387,399,400,401,408,409, 413,432,435,444,451,459
stem and leaf method(茎叶方法),270
step function(阶梯函数),52
stratified samples(分层样本),362
sum of chi-squared random variables(χ^2 随机变量的和),62
sum of integers formula(整数和公式),40
sum of random variables(随机变量的和):
mean of(均值),39
variance of(方差),48
sum of ranks(秩和):
mean of(均值),41,49
variance of(方差),48,49
sum of squared integers formula(整数平方和公式),43
sure event(必然事件),14
survival curves(生存曲线),119
survival function(生存函数),89
empirical(经验),89
symmetric distributions(对称分布),350,351
symmetry, Smirnov test for(Smirnov 对称性检验),465
symmetry, tests for(对称性检验),364
SYSTAT,88,91,259

T

t distribution, table(t 分布,表格),561
t statistic computed on ranks(基于秩计算的 t 统计量),367
t test(t 检验):
efficiency of paired(配对效率),364
one sample(一样本),363,418
paired(配对),363
two sample(两样本),284,417
table, contingency(列联表),166,179,292
target population(目标总体),69,70
tau, Kendall's (τ, Kendall), 318, 319, 325, 326,335

test, conservative(保守检验), 113
test hypothesis(假设检验), 95
test, one tailed(单边检验), 98
test statistic(检验,统计量), 35, 96, 97
test, two tailed(双边检验), 98
test, unbiased(无偏检验), 106, 108, 160
testing hypotheses(假设检验), 95
tests, consistent sequence of(相合序列检验), 106, 108, 160
three-way contingency table(三种方式列联表), 214
tolerance limits(容忍限), 150
  approximation formulas for(逼近公式), 151, 155
  exact tables for(精确表), 537
transformation, rank(秩变换), 417
trend(趋势):
  Cox and Stuart test for(Cox 和 Stuart 检验), 169, 170
  Daniel's test for(Daniel 检验), 323
trials(基本试验), 6
Tschuprow's coefficient(Tschuprow 系数), 232
two independent samples, randomization test for(两独立样本的随机化检验), 409
two-sample Cramér-von Mises test(两样本 Cramér-von Mises 检验), 463
two-sample t test(两样本 t 检验), 198
two-tailed test(双边 t 检验), 98
two-way contingency table(两种方式列联表), 214
type Ⅰ error(一类错误), 98
type Ⅱ error(二类错误), 98, 99

U

unbiased estimator(无偏估计量), 84, 94
unbiased, sign test(无偏,符号检验), 163
unbiased test(无偏检验), 106, 108, 160
uniform distribution(均匀分布):
  continuous(连续), 433
  discrete(离散), 28, 437
upper-tailed test(右边检验), 98

V

value, expected(期望值), 35, 39
van der Waerden test(van der Waerden 检验), 397
variable, random(随机变量), 22, 23, 76
variance(方差), 36, 37
  in binomial distribution(二项分布), 49
  multiple comparisons for test for(检验的多重比较), 304
  in rank test using scores(得分秩检验), 307
  sample(样本), 81, 83
  squared ranks test for(平方秩检验), 300
  of sum of random variables(随机变量的和), 48
  of sum of ranks(秩和), 48, 49
  tests for(检验), 309
variations of the sign test(符号检验的变差), 166, 175

W

Walsh test(Walsh 检验), 364
Wilcoxon signed ranks test(Wilcoxon 符号秩检验), 164, 352, 411
  continuity correction in(连续相关), 359
  efficiency of(效率), 364
  extension to k samples of(推广到 k 个样本), 367
  normal approximation in(正态逼近), 353, 359
  tables(表格), 547-548
Wilcoxon test(Wilcoxon 检验), 103
Wilcoxon two-sample test(Wilcoxon 两样本检验), 271

［Ｇｅｎｅｒａｌ Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝实用非参数统计 （第３版）
作者＝（美）Ｗ．Ｊ．Ｃｏｎｏｖｅｒ著；崔恒建译
页数＝４１４
出版社＝北京市：人民邮电出版社
出版日期＝２００６．０４
ＳＳ号＝１１５８４１００
ＤＸ号＝０００００４７０５２６０
ＵＲＬ＝ｈｔｔｐ：／／ｂｏｏｋ．ｓｚｄｎｅｔ．ｏｒｇ．ｃｎ／ｂｏｏｋＤｅｔａｉｌ．ｊｓｐ？ｄｘＮｕｍｂｅｒ＝０００００４７０５２６０＆ｄ＝Ｅ７３２ＣＦ３Ｆ４６６Ｆ３５０２Ｂ６９３Ｂ３３６４Ｆ０Ｅ８４５５

封面
书名
版权
前言
目录
第１章　概率论
导言
１．１　计数
１．２　概率
１．３　随机变量
１．４　随机变量的性质
１．５　连续型随机变量
１．６　第１章复习题
第２章　统计推断
导言
２．１　总体、样本与统计量
２．２　估计
２．３　假设检验
２．４　假设检验的性质
２．５　非参数统计评述
２．６　第２章复习题
第３章　基于二项分布的检验
导言
３．１　二项检验与ｐ的估计
３．２　分位数检验和ｘｐ的估计
３．３　容忍限
３．４　符号检验
３．５　符号检验的一些变形
第４章　列联表
导言
４．１　２×２列联表
４．２　ｒ×ｃ列联表
４．３　中位数检验
４．４　相依性度量
４．５　ｘ２拟合优度检验
４．６　相关观测的Ｃｏｃｈｒａｎ检验
４．７　其他分析方法讨论
４．８　第３、４章复习题
第５章　秩检验
导言
５．１　两个独立样本
５．２　多个独立样本
５．３　等方差检验
５．４　秩相关性度量
５．５　非参数线性回归方法
５．６　单调回归方法
５．７　一样本或配对情形
５．８　多个相关的样本
５．９　平衡的不完全区组设计
５．１０　Ａ．Ｒ．Ｅ．不低于１的检验
５．１１　Ｆｉｓｈｅｒ随机化方法
５．１２　秩变换的讨论
５．１３　第５章复习题
第６章　Ｋｏｌｍｏｇｏｒｏｖ－Ｓｍｉｒｎｏｖ型统计量

导言
６．１　Ｋｏｌｍｏｇｏｒｏｖ拟合优度检验
６．２　分布族的拟合优度检验
６．３　两组独立样本的检验
６．４　第１章至第６章复习题
附表
奇数号习题解答
索引
参考文献